

81929

From: Hutzell, Paula
Sent: Monday, December 09, 2002 1:08 PM
To: STIC-Biotech/ChemLib; Graser, Jennifer
Subject: FW: rush search

approved

-----Original Message-----

From: Graser, Jennifer
Sent: Monday, December 09, 2002 12:28 PM
To: Hutzell, Paula
Subject: rush search

Hi Paula,

- Could you please authorize the following rush search for a **RCE** which is due?
Thanks,
Jennifer

STIC:

Please search amino acids 1-40 from SEQ ID NOS:1-5 from 09/142,970 in pending and commercial databases.

Thanks,
Jennifer Graser
CM1 7E09
AU 1645
308-1742

Edward Hart
Technical Info. Specialist
STIC/Biotech
CMI 6B02 Tel: 305-9203

only
get hits
from full-length
TGA
none of the
ref. tech consisting
1-40.

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 12/10/02
Date Completed: 12/10/02
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:

NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)

STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

THIS PAGE BLANK (USPTO)

58 2002

us-09-142-970-1_copy_1_40.rag

Page 1

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:14:59 : Search time 28.2 seconds
(without alignments)
189,008 Million cell updates/sec

Title: US-09-142-970-1_COPY_1_40

Perfect score: 220
Sequence: 1 LYYKNRYALKSGSVNAPMPENGOTENDMILMGSTQE 40

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	220	100.0	105	19	AAW65656
2	220	100.0	105	19	AAW61602
3	203	92.3	104	19	AAW61604
4	202	91.8	104	19	AAW61603
5	190	86.4	104	19	AAW61606
6	189	85.9	104	19	AAW61605
7	98	44.5	154	11	AAW61605
8	59	26.8	359	22	AAW61605
9	57	25.9	452	22	AAW61605
10	57	25.9	452	22	AAW61605

11	56	25.5	452	22	AAU34434
12	56	25.5	452	22	AAU34434
13	55	25.0	452	22	AAU34434
14	54.5	24.8	948	22	AAU34434
15	53.5	24.3	236	22	AAU34434
16	53.5	24.3	236	22	AAU34434
17	53.5	24.3	236	22	AAU34434
18	53.5	24.3	236	22	AAU34434
19	53	24.1	479	22	AAU34434
20	53	24.1	569	19	AAU34434
21	53	24.1	569	19	AAU34434
22	53	24.1	591	19	AAU34434
23	53	24.1	591	19	AAU34434
24	53	24.1	678	22	AAU34434
25	53	24.1	845	22	AAU34434
26	53	24.1	845	22	AAU34434
27	52.5	23.9	194	22	AAU34434
28	52	23.6	392	22	AAU34434
29	52	23.6	392	22	AAU34434
30	52	23.6	392	22	AAU34434
31	52	23.6	392	22	AAU34434
32	52	23.6	392	22	AAU34434
33	52	23.6	392	22	AAU34434
34	52	23.6	641	15	AAU34434
35	52	23.6	954	21	AAU34434
36	51.5	23.4	862	22	AAU34434
37	51.5	23.4	862	22	AAU34434
38	51.5	23.4	895	22	AAU34434
39	51	23.2	374	23	AAU34434
40	51	23.2	411	22	AAU34434
41	51	23.2	411	22	AAU34434
42	51	23.2	1319	14	AAU34434
43	50.5	23.0	317	23	AAU34434
44	50.5	23.0	394	22	AAU34434
45	49.5	22.5	230	21	AAU34434

ALIGNMENTS

RESULT 1	AAW65656 standard; peptide: 105 AA.
ID	AAW65656
AC	AAW65656
DT	15-OCT-1998 (first entry)
DE	105-mer peptide used in polysaccharide-peptide conjugate.
XX	Polysaccharide-peptide conjugate; vaccine; linker moiety; adjuvant; immune response.
KW	Synthetic.
OS	Synthetic.
FT	Key
FT	Modified-site
XX	Location/Qualifiers
XX	1 /note="N-terminal acetyl"
XX	23-JUL-1998.
XX	21-JAN-1998; 98MO-EP00654.
XX	21-JAN-1997; 97EP-0100884.
XX	(INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX	Mistretta N, Moreau M;
XX	WPI, 1998-413820/35.
XX	

E. coli cellular p
Escherichia coli p
Modified carboxyl
Drosophila melanog
C. glutamicum prote
Coriobacterium gl
Pseudomonas aerugi
Escherichia coli p
Novel human diago
Streptococcus pneu
S. pneumoniae SP08
S. pneumoniae SP04
Streptococcus pneu
Novel human diago
Novel human diago
Human ORFX protein
Human immune/haema
Arabidopsis thaliana
Abscisic acid resp
A. thaliana abscis
Abscisic acid resp
Isomaltodextranase
Rumicoccus flavo
Human protein SEQ
Novel human dynami
Human protein SEQ
Herbicideally activ
Drosophila melanog
Cellulase AE-1. A
Cellulose synthase
Arabidopsis transc
Drosophila melanog
Human ORFX ORF21 p

PT Polysaccharide-peptide conjugates, useful as, e.g. vaccines -
 PT comprise peptide moiety with at least six amino acid residues,
 PT polysaccharide chain with at least four repeat units, and linker
 XX moiety

PS Example 1; Page 14; 28pp; English.

CC The invention relates to: (A) polysaccharide-peptide conjugate (in which
 CC the polysaccharide is immunogenic), comprising: (a) a peptide moiety
 CC which has at least 6 amino acid residues, at least 1 of which is a
 CC cysteine residue; (b) a polysaccharide chain comprising at least 4 repeat
 CC units; and (c) a linker moiety bound to the thiol group of the cysteine.
 CC The linker is also bound to: (i) native amino, hydroxyl or carboxyl
 CC groups of the polysaccharide chain; (ii) amino groups created by
 CC hydrolysis of native N-acetyl groups of the polysaccharide chain, or
 CC derivatisation with a spacer moiety bound to the polysaccharide chain upon
 CC carboxyl groups of the polysaccharide chain, and (B) conjugating a
 CC peptide as in (Aa) to a polysaccharide chain comprising at least 4 repeat
 CC units, comprising: (a) coupling the peptide to a linker through the thiol
 CC group of the cysteine residue, and (b) coupling the linker to the thiol
 CC described in (A). The conjugates are especially useful as vaccines to
 CC elicit a protective long term immune response against a pathogenic
 CC microorganism from which the immunogenic polysaccharide chain is derived.
 CC known polysaccharide-peptide conjugates are less immunogenic than their
 CC polysaccharide-protein counterparts and require adjuvantation. The new
 CC conjugates have good immunogenicity, largely because the peptides are of
 CC sufficient size. The present sequence represents a peptide used in
 CC a polysaccharide-peptide conjugate.

Sequence 105 AA:

Query Match 100.0%; Score 220; DB 19; Length 105;
 Best Local Similarity 100.0%; Pred. No. 9.5e-24;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYYKNRYVALKSGSVNAPMPENGOTENDWILMGSTOE 40
 DB 2 LYYKNRYVALKSGSVNAPMPENGOTENDWILMGSTOE 41

RESULT 2
 AAW61602
 ID AAW61602 standard; peptide: 105 AA.
 AC AAW61602;
 XX
 DT 27-OCT-1998 (first entry)
 XX
 DE Neisseria IgA1 protease fragment 1.
 XX
 KW Immunoglobulin protease; carrier; paediatric; vaccine;
 XX epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
 OS Neisseria sp.
 XX
 PN W09831791-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 20-JAN-1998; 98WO-EP00294.
 XX
 PR 21-JAN-1997; 97EP-0100883.
 XX

PA (PLAC) MAX PLANCK GES. FÖRDERUNG WISSENSCHAFTEN.
 XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
 XX
 PI Achtmann M, Moreau M;
 XX
 DR WPI: 1998-414092/35.
 XX

New peptide from Neisseria immunoglobulin protease - useful as

applied

PT immunogenic carrier, e.g. particularly for polysaccharide(s),
 PT forming conjugates used in vaccines against Neisseria and
 PT Haemophilus
 XX
 XX

PS Claim 6; Page 10; 32pp; English.

CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.

Sequence 105 AA:

Query Match 100.0%; Score 220; DB 19; Length 105;
 Best Local Similarity 100.0%; Pred. No. 9.5e-24;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYYKNRYVALKSGSVNAPMPENGOTENDWILMGSTOE 40
 DB 2 LYYKNRYVALKSGSVNAPMPENGOTENDWILMGSTOE 41

RESULT 3
 AAW61604
 ID AAW61604 standard; peptide: 104 AA.
 AC AAW61604;
 XX
 DT 27-OCT-1998 (first entry)
 XX
 DE Neisseria IgA1 protease fragment 3.
 XX
 KW Immunoglobulin protease; carrier; paediatric; vaccine;
 XX epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
 OS Neisseria sp.
 XX
 PN W09831791-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 20-JAN-1998; 98WO-EP00294.
 XX
 PR 21-JAN-1997; 97EP-0100883.
 XX

PA (PLAC) MAX PLANCK GES. FÖRDERUNG WISSENSCHAFTEN.
 XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
 XX
 PI Achtmann M, Moreau M;
 XX
 DR WPI: 1998-414092/35.
 XX
 PD 23-JUL-1998.
 XX
 PF 20-JAN-1998; 98WO-EP00294.
 XX
 PR 21-JAN-1997; 97EP-0100883.
 XX

PT New peptide from Neisseria immunoglobulin protease - useful as
 PT immunogenic carrier, e.g. particularly for polysaccharide(s),
 PT forming conjugates used in vaccines against Neisseria and
 PT Haemophilus
 XX
 XX

PS Claim 2; Fig 1; 32pp; English.
 CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.

Sequence 104 AA:

XX (PIAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
 XX Achtmann M, Moreau M;
 PI WPI: 1998-414092/35.
 XX
 DR New peptide from Neisseria immunoglobulin protease - useful as
 PT immunogenic carrier, e.g., particularly for polysaccharide(s),
 PT forming conjugates used in vaccines against Neisseria and
 PT Haemophilus
 PS Claim 2: Fig 1/4: 32pp: English.
 XX
 CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.
 CC
 CC Sequence 104 AA:
 SQ
 Query Match 85.9%; Score 189; DB 19; Length 104;
 Best Local Similarity 82.5%; Pred. NO. 2.6e-19;
 Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
 QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
 DB 1 LYKKNRYVALKSGSVNAPMPENGVAENNDWFMGCTOE 40
 RESULT 7
 AAR07304
 ID AAR07304 standard; protein: 1541 AA.
 AC AAR07304:
 XX
 DT 31-JAN-1991 (first entry)
 XX
 DE Igal protease.
 XX
 KW Igal: vaccine: meningitis; gonorrhoea; allergies.
 XX
 OS Haemophilus influenzae.
 XX
 PN WO9011367-A.
 XX
 PD 04-OCT-1990.
 XX
 PF 16-MAR-1990; 90WO-DK00073.
 XX
 PR 17-MAR-1989; 89DK-0001308.
 XX
 PA (KILI/) KILIAN M.
 XX
 PI Kilian M, Poulsen K;
 XX
 DR WPI: 1990-320267/42.
 DR N-PSDB: AAO06164.
 XX
 PT Immunoglobulin A1 protease prodn. - by cloning from
 PT microorganisms for immunisation against immunoglobulin A1
 PT protease producing bacteria
 XX
 PS Disclosure: fig 3; 44pp: English.
 CC This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
 CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,
 CC gonorrhoea or allergic diseases. It specifically cleaves the heavy

for cons

CC chain of human Igal in the hinge region.
 XX
 SQ Sequence 1541 AA:
 Query Match 44.5%; Score 98; DB 11; Length 1541;
 Best Local Similarity 40.08%; Pred. NO. 7.9e-05;
 Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;
 QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
 DB 595 LNIENYTYALKSGSTSELPKNSGESNENWLYMGKTSO 634
 RESULT 8
 AAB51254
 ID AAB51254 standard; protein: 359 AA.
 AC AAB51254:
 XX
 DT 27-MAR-2001 (first entry)
 XX
 DE Human CD72 protein sequence SEQ ID NO:7.
 XX
 KW CD100; CD72: screening: viral infection; bacterial infection; cancer;
 KW fungal infection; infectious disease; allergy; autoimmune disease.
 XX
 OS Homo sapiens.
 XX
 PN WO200075655-A1.
 XX
 PD 14-DEC-2000.
 XX
 PF 01-JUN-2000; 2000WO-JP03558.
 XX
 PR 03-JUN-1999; 99JP-0157111.
 XX
 PA (TAKE) TAKEDA CHEM IND LTD.
 XX
 PI Kikutan H, Kumonogoh A, Hori A;
 XX
 DR WPI: 2001-061765/07.
 DR N-PSDB: AAB51254.
 XX
 PT Screening of compounds modifying the binding of CD100 to CD72 for use
 PT in treatment of infectious diseases, cancer and disorders of antibody
 PT production
 XX
 PS Disclosure: Page 105-106; 110pp: Japanese.
 XX
 CC The present invention describes a method of screening for compounds
 CC which modify the binding of CD100 and its salts to CD72 and its salts,
 CC which modify the binding of CD100 and its salts to CD72 and its salts,
 CC CD100 is contacted with CD72 in the presence and absence of the compound
 CC and the degree of binding is compared. Also described are: (1) kits for
 CC carrying out the novel method; (2) compounds identified by (2);
 CC method; (3) drug compositions containing the compounds of (2);
 CC (4) non-human CD100 knockout animals; (5) screening compounds as drugs
 CC for the prevention and treatment of CD100-associated diseases, using the
 CC knockout animals of (4); (6) transgenic non-human animals transformed
 CC with a CD100 gene or modified CD100 gene; and (7) screening methods
 CC using the transgenic animals of (6). The method can be used for
 CC identifying compounds for the treatment and prevention of bacterial,
 CC viral and fungal infections, cancers, and diseases involving abnormal
 CC antibody production or excess antibody production, such as allergies and
 CC autoimmune diseases. The present sequence represents human CD72 which is
 CC given in the exemplification of the present invention.
 XX
 SQ Sequence 359 AA:
 Query Match 26.8%; Score 59; DB 22; Length 359;
 Best Local Similarity 37.2%; Pred. NO. 5.4;
 Matches 16; Conservative 6; Mismatches 15; Indels 6; Gaps 2;
 QY 1 LYKKNRYVALKS---GGSVNAPMPENGOTENNNDWILMGSTO 39

Db 274 IYPSHSHYPLNSILPNGSGNSYWM--TGLSNKMDKLTDDTQ 314

RESULT 9

AAU38481 standard; Protein: 1364 AA.

AAU38481;

27-JUL-2001 (first entry)

S cerevisiae apoptosis associated protein YGR183C.

Yeast: fungus; apoptosis; infection; proliferative disease; vaccine; autoimmune disease; ischaemia; neurodegeneration.

Saccharomyces cerevisiae.

WO200102550-A2.

11-JAN-2001.

03-JUL-2000; 2000WO-BE00077.

01-JUL-1999; 99EP-0870141.

(JANC) JANSSEN PHARM NV.

Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;

Nelissen BJM, Reekmans RJ;

WPI: 2001-367042/38.

N-PSDB; AAH29948.

Yeast and fungal nucleic acids encoding proteins involved in a pathway leading to programmed cell death, useful for treating proliferative disorders, yeast and fungal infections, or for preventing apoptosis in certain diseases -

Claim 1; Fig 1; 218pp; English.

The present invention provides the protein and coding sequences of a number of apoptosis associated proteins from the yeast *Saccharomyces cerevisiae* and the fungus *Candida albicans*. These can be used to identify treatments for fungal and yeast infections, for proliferative diseases and for apoptosis related diseases such as autoimmune diseases, ischaemia and neurodegeneration. The present sequence is one of the *S. cerevisiae* proteins of the invention.

Sequence 1364 AA;

Query Match 26.1%; Score 57.5; DB 22; Length 1364;

Best Local Similarity 37.9%; Pred. No. 44; Mismatches 11; Conservative 6; Mismatches 7; Indels 5; Gaps 1;

OY 3 YKRYRYALKSGSVNAPMPCOTENND 31

Db 519 YRSHREISIKSGSVG-----NSNTGND 542

RESULT 10

AAU38481 standard; Protein: 452 AA.

AAU38481;

14-FEB-2002 (first entry)

Salmonella typhi cellular proliferation protein #372.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

OS *Salmonella typhi*.

WO200170955-A2.

27-SEP-2001.

21-MAR-2001; 2001WO-US09180.

21-MAR-2000; 2000US-191078P.

23-MAY-2000; 2000US-206848P.

26-MAY-2000; 2000US-207727P.

23-OCT-2000; 2000US-242578P.

27-NOV-2000; 2000US-253625P.

22-DEC-2000; 2000US-257931P.

16-FEB-2001; 2001US-269308P.

(ELIT-) ELITRA PHARM INC.

Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

Yamamoto RT, Xu HH;

WPI: 2001-611495/70.

N-PSDB; AAS56340.

New polynucleotides for the identification and development of

antibiotics, comprise sequences of antisense nucleic acids -

Example 3; Seq ID No 14074; 511pp; English.

The invention relates to antisense inhibitors of genes essential to prokaryotic cellular proliferation, their use in identifying the genes themselves and the encoded proteins. The prokaryotes used are *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also useful for the identification of potential new targets for antibiotic development. The antisense nucleic acids can also be used to identify proteins used in proliferation, to express these proteins, and to obtain antibodies capable of binding to the expressed proteins. The proteins can be used to screen compounds in rational drug discovery programmes. The antisense nucleic acid sequence is also useful to screen for homologous nucleic acids which are required for cell proliferation in a wide variety of organisms. The present sequence represents an essential prokaryotic cellular proliferation protein. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 452 AA;

Query Match 25.9%; Score 57; DB 22; Length 452;

Best Local Similarity 33.3%; Pred. No. 14; Mismatches 12; Conservative 7; Mismatches 9; Indels 8; Gaps 2;

OY 12 KSGGSVNAKPPENG-----QTEENDWT---LMGSTQ 39

Db 200 KAKGEITYGLPENGIAIMNADNNDMLNMOISIIDRQ 235

RESULT 11

AAU34434 standard; Protein: 452 AA.

AAU34434;

14-FEB-2002 (first entry)

E. coli cellular proliferation protein #15.

Antisense; prokaryotic cellular proliferation protein;

antibiotic; antibacterial; drug design.

```

XX OS Escherichia coli.
XX XX
XX PN WO200170955-A2.
XX PD
XX PF 27-SEP-2001.
XX PE 21-MAR-2001; 2001WO-US09180.
XX PR 21-MAR-2000; 2000US-191078P.
XX PR 23-MAY-2000; 2000US-206848P.
XX PR 26-MAY-2000; 2000US-207727P.
XX PR 23-OCT-2000; 2000US-242578P.
XX PR 27-NOV-2000; 2000US-253625P.
XX PR 22-DEC-2000; 2000US-257931P.
XX PR 16-FEB-2001; 2001US-269308P.
XX PA (ELIT-) ELITRA PHARM INC.
XX PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX PI Yamamoto RT, Xu HH;
XX DR N-PSDB; AAS52293.
XX PS WPI: 2001-611495/70.
XX PT New polynucleotides for the identification and development of
XX PT antibiotics, comprise sequences of antisense nucleic acids -
XX PS Example 3; Seq ID No 10027; 511pp; English.
XX CC The invention relates to antisense inhibitors of genes essential to
XX CC prokaryotic cellular proliferation, their use in identifying the
XX CC genes, their use in the discovery of novel antibiotics, the essential
XX CC genes themselves and the encoded proteins. The prokaryotes used are
XX CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
XX CC pneumoniae, pseudomonas aeruginosa and Enterococcus faecalis. The
XX CC invention is also useful for the identification of potential new targets
XX CC for antibiotic development. The antisense nucleic acids can also be used
XX CC to identify proteins used in proliferation, to express these proteins,
XX CC and to obtain antibodies capable of binding to the expressed protein.
XX CC The proteins can be used to screen compounds in rational drug discovery
XX CC programmes. The antisense nucleic acid sequence is also useful to screen
XX CC for homologous nucleic acids which are required for cell proliferation in
XX CC a wide variety of organisms. The present sequence represents an
XX CC essential prokaryotic cellular proliferation protein.
XX CC Note: The sequence data for this patent did not form part
XX CC of the printed specification, but was obtained in electronic
XX CC format directly from WIPO at
XX CC ftp.wipo.int/pub/published_pct_sequences.
XX SC Sequence 452 AA:
SY Query Match 25.5%; Score 56; DB 22; Length 452;
Best Local Similarity 38.5%; Pred. No. 19;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1.
QY 12 KSGSVNAPMPENG---QTENNMMT 33
Db 200 KAKGETFGLPENGIAIMADNNMDL 225
| : | : : |||||
RESULT 12
AAG98402
ID AAG98402 standard; Protein; 452 AA.
XX AC AAG98402;
XX DT 21-SEP-2001 (first entry)
DE Escherichia coli protein sequence SEQ ID NO:450.
XX ES Escherichia coli; identification; proliferation; microorganism;
KW antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
```

XX	bacterial growth inhibition.
OS	Escherichia coli.
PN	MO200148209-A2.
XX	
PD	05-JUL-2001.
XX	
PE	19-DEC-2000; 2000MO-US34419.
XX	
PR	23-DEC-1999; 9905-0173005.
XX	
PA	(ELIT-) ELITRA PHARM INC.
XX	
PI	Forsyth RA, Ohlsen KL, Zyskind JW;
XX	
DR	WPI. 2001-457376/49.
XX	
DR	N-PSDB; AAH81458.
XX	
PT	Novel nucleic acids encoding proteins required for Escherichia coli
XX	proliferation, useful for screening for antimicrobial agents -
XX	
PS	Claim 19; Page 569; 596pp; English.
XX	
CC	The present invention describes a purified or isolated nucleic acid
CC	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC	microorganism is capable of inhibiting proliferation of a microorganism.
CC	(I) have antibacterial and antibiotic activities, and can be used in
CC	gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC	of the microorganism, and the manufactured antibiotic is useful for
CC	reducing the activity or level of a gene product required for
CC	proliferation of a microorganism in a subject, specifically humans. The
CC	nucleic acids that inhibit bacterial growth or proliferation can be used
CC	as antisense therapeutics for killing bacteria. In addition to
CC	therapeutic applications, the nucleic acid sequences complementary to
CC	sequences required for proliferation can be used as diagnostic tools.
CC	For example, nucleic acid probes complementary to proliferation-required
CC	sequences that are specific for particular species of microorganisms can
CC	be used as probes to identify particular microorganism species in
CC	clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC	proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC	represent oligonucleotides, which are used in the exemplification of the
CC	present invention.
XX	
SQ	Sequence 452 AA;
XX	
Query Match	25.5%; Score 56; DB 22; Length 452;
Best Local Similarity	38.5%; Pred. NO. 19;
Matches 10; Conservative	6; Mismatches 6; Indels 4; Gaps 1
OY	12 KSGGSVNAHPENG---OTENDMT 33
	I : I : : : I I I I : I I I I :
Db	200 KAKGEIFSGIPENGIAIMNADNNDWL 225
RESULT 13	
AAAR07532	
ID	AAAR07532 standard; protein; 300 AA.
XX	
XX	AAAR07532;
XX	
DT	01-FEB-1991 (first entry)
XX	
DE	Modified carboxyl esterase for hydrolysis of S-naproxen.
XX	
KW	Carboxyl esterase; hydrolysis; S-naproxen.
XX	
OS	Bacillus subtilis Thai 1-8.
XX	
FH	Key Location/Qualifiers
FT	Misc-difference 34..34
FT	/label=E, R

2002

us-09-142-970-1-copy_1_40.rapb

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

protein - protein search, using sw model
December 10, 2002, 10:51:54 ; Search time 5.2 Seconds
(without alignments)
124.941 Million cell updates/sec

Run on:

Title: US-09-142-970-1_COPY_1_40
Perfect score: 220
Sequence: 1 LYKKNRYALKSGSVNAPPEKGTENNMDWLMGSGR 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 103943 seqs, 16242309 residues
Total number of hits satisfying chosen parameters: 103943

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

1:	/cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2:	/cgn2_6/ptodata/1/pubpaa/PCRT_NEW_PUB.pep:*
3:	/cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4:	/cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
5:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6:	/cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7:	/cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8:	/cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
9:	/cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10:	/cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
11:	/cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
12:	/cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
13:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
14:	/cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

pred. No. is the number of results predicted by the change to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	57	25.9	452	10	US-09-815-242-14074
2	56	25.5	452	10	US-09-741-669-450
3	56	25.5	452	10	US-09-815-242-10027
4	54.5	24.8	421	10	US-09-420-785A-4
5	54.5	24.8	421	10	US-09-901-252-15
6	53.5	24.3	1317	10	US-09-815-242-5118
7	53	24.1	569	10	US-09-765-272-154
8	53	24.1	591	10	US-09-765-272-74
9	51.5	23.4	863	10	US-09-796-100-2
10	49.5	22.5	235	10	US-09-947-442-2
11	48	21.8	275	10	US-09-846-808-8
12	48	21.8	275	10	US-10-059-964-28
13	48	21.8	396	10	US-09-740-026A-4
14	48	21.8	1207	9	US-10-108-605-71
15	48	21.8	3084	10	US-09-938-275-4
16	47.5	21.6	281	12	US-10-059-964-58
17	47.5	21.6	348	10	US-09-903-799-2
18	47.5	21.6	730	10	US-09-780-053-2
19	47	21.4	419	10	US-09-815-242-12095

20	46.5	21.1	791	10	US-09-804-472-2	Sequence 2, Appl1
21	46.5	21.1	844	9	US-10-027-806-30	Sequence 30, Appl1
22	46.5	21.1	845	9	US-10-027-806-62	Sequence 62, Appl1
23	46	20.9	73	10	US-09-771-161A-155	Sequence 155, Appl
24	46	20.9	171	10	US-09-764-864-1176	Sequence 1176, Ap
25	46	20.9	191	10	US-09-764-864-1584	Sequence 1584, Ap
26	46	20.9	213	10	US-09-771-161A-156	Sequence 246, App
27	46	20.9	479	10	US-09-771-161A-246	Sequence 247, App
28	46	20.9	479	10	US-09-771-161A-247	Sequence 248, App
29	46	20.9	915	10	US-09-771-161A-248	Sequence 2, Appl1
30	46	20.9	3075	10	US-09-332-226-2	Sequence 176, App
31	46	20.7	73	10	US-09-938-275-5	Sequence 129, App
32	45.5	20.5	296	10	US-09-764-846-176	Sequence 72, Appl
33	45	20.5	299	10	US-09-960-472-1	Sequence 86, Appl
34	45	20.5	331	10	US-09-820-893-72	Sequence 62, Appl
35	45	20.5	331	10	US-09-820-893-129	Sequence 5, Appl1
36	45	20.5	1007	9	US-09-843-676-86	Sequence 2, Appl1
37	45	20.5	1031	9	US-09-843-676-2	Sequence 62, Appl
38	44.5	20.2	247	9	US-09-991-496-62	Sequence 62, Appl
39	44.5	20.2	247	10	US-09-874-923-62	Sequence 5, Appl1
40	44.5	20.2	295	10	US-09-846-808-5	Sequence 22, Appl
41	44.5	20.2	295	12	US-10-059-964-22	Sequence 5912, Ap
42	44.5	20.2	331	10	US-09-815-242-5912	Sequence 12872, A
43	44.5	20.2	331	10	US-09-815-242-12872	Sequence 13146, A
44	44.5	20.2	331	10	US-09-815-242-13146	Sequence 243, App
45	44.5	20.2	876	9	US-09-712-363-243	

ALIGNMENTS

RESULT 1
US-09-815-242-14074
Sequence 14074, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Tralick, John D.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 14074
LENGTH: 452
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(452)
OTHER INFORMATION: Xaa = Any Amino Acid

US-09-815-242-14074

Query Match

Best Local Similarity 25.5%; Score 57; DB 10; Length 452;
Matches 12; Conservative 7; Mismatches 9; Indels 8; Gaps 2;

OY 12 KSGGSVNAHPENG---QTEENNMI 33
DB 200 KAKGEIFSGLPENGIAIMNADNDMLNMSITIGDRQ 235

RESULT 2

US-09-741-669-450
; Sequence 450, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009a
; CURRENT APPLICATION NUMBER: US/09/741,669
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; NUMBER OF SEQ ID NOS: 1999-12-23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 450
; LENGTH: 452
; TYPE: PRF
; ORGANISM: Escherichia coli
US-09-741-669-450

Query Match

Best Local Similarity 25.5%; Score 56; DB 10; Length 452;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY 12 KSGGSVNAHPENG---QTEENNMI 33
DB 200 KAKGEIFSGLPENGIAIMNADNDML 225

RESULT 3

US-09-815-242-10027
; Sequence 10027, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Traawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22

PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16

NUMBER OF SEQ ID NOS: 14110

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 10027

LENGTH: 452

TYPE: PRF

ORGANISM: Escherichia coli

US-09-815-242-10027

Query Match 25.5%; Score 56; DB 10; Length 452;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY 12 KSGGSVNAHPENG---QTEENNMI 33
DB 200 KAKGEIFSGLPENGIAIMNADNDML 225

RESULT 4

US-09-420-785A-4
; Sequence 4, Application US/09420785A
; Patent No. US20010010923A1
; GENERAL INFORMATION:
; APPLICANT: MORTENSEN, UFFE
; APPLICANT: OLESEN, KUTBLD
; APPLICANT: STENNICKE, HENNING
; APPLICANT: SORENSEN, STEEN B.
; APPLICANT: BREDDAM, KLAUS
; TITLE OF INVENTION: MODIFIED CARBOXYPEPTIDASE
; FILE REFERENCE: 089187/0109
; CURRENT APPLICATION NUMBER: US/09/420,785A
; PRIOR FILING DATE: 1999-10-19
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 421
; TYPE: PRF
; ORGANISM: Saccharomyces cerevisiae
US-09-420-785A-4

Query Match

Best Local Similarity 24.8%; Score 54.5; DB 10; Length 421;
Matches 12; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

OY 3 YKNRYRYALKSG-SYNADHPENGQTEENNMI 33
DB 384 YKHETTYIRVFNHGHWYFDPVFNALSMYVMEWI 415

RESULT 5

US-09-901-252-15
; Sequence 15, Application US/09901252
; Patent No. US2002002658A1
; GENERAL INFORMATION:
; APPLICANT: Chapple, Clint
; TITLE OF INVENTION: Genes Encoding Sinapoylglucose:Malate Sinapoyltransferase and
; FILE REFERENCE: N1422-004
; CURRENT APPLICATION NUMBER: US/09/901,252
; PRIOR FILING DATE: 2001-07-09
; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 421
; TYPE: PRF
; ORGANISM: Saccharomyces
US-09-901-252-15

Query Match

Best Local Similarity 24.8%; Score 54.5; DB 10; Length 421;
Matches 12; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-765-272-74

Query Match 24.1%; Score 53; DB 10; Length 591;
Best Local Similarity 38.7%; Pred. No. 14;
Matches 12; Conservative 4; Mismatches 5; Indels 10; Gaps 1;

OY 3 YKRYRYALKSGSVNAPMPENGOTENNMI 33
DB 190 FENGHYLYLKSGGYMA-----NEMI 210

RESULT 9
US-09-796-100-2
Sequence 2, Application US/09796100
Patent No. US20020076784A1

GENERAL INFORMATION:
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 40322, A No. US20020076784A1 Human Dynamin
FILE REFERENCE: 35800/209283
CURRENT APPLICATION NUMBER: US/09/796,100
CURRENT FILING DATE: 2001-02-28
PRIOR FILING DATE: 2000-02-28
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 863
TYPE: PRT
ORGANISM: Homo sapiens
US-09-796-100-2

Query Match 23.4%; Score 51.5; DB 10; Length 863;
Best Local Similarity 30.0%; Pred. No. 36;
Matches 15; Conservative 3; Mismatches 9; Indels 23; Gaps 1;

OY 3 YKRYRY-----YALKSGSVNAPMPENGOTEN 29
DB 593 YKDYRFLEACDSQEDVDYMKASILLRAGVYPPDKSVGNKKAENDENGOAEN 642

RESULT 10
US-09-947-442-2

Sequence 2, Application US/09947442
Patent No. US20020052486A1
GENERAL INFORMATION:
APPLICANT: BATHÉ, BRIGITTE
APPLICANT: SCHROEDER, INDRÀ
APPLICANT: PFEFFERLE, WALTER
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES WHICH CODE FOR THE GPMB GENE
FILE REFERENCE: 213067UD50X
CURRENT APPLICATION NUMBER: US/09/947,442
CURRENT FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: DE 10044772.4
PRIOR FILING DATE: 2000-09-09
PRIOR APPLICATION NUMBER: DE 10133668.3
PRIOR FILING DATE: 2001-07-11
NUMBER OF SEQ ID NOS: 4
SOFTWARE: PatentIn version 3.1
SEQ ID NO 2
LENGTH: 235
TYPE: PRT
ORGANISM: Corynebacterium glutamicum
US-09-947-442-2

Query Match 22.5%; Score 49.5; DB 10; Length 235;
Best Local Similarity 38.5%; Pred. No. 14;
Matches 15; Conservative 2; Mismatches 13; Indels 9; Gaps 2;

OY 7 RRYA-----LKSGSV-NAPMPENGOTENNMI 36
DB 191 RRYAGSENPEDDLKISSAVSNPHEGNNVENAOWLWG 229

RESULT 11
US-09-846-808-8
Sequence 8, Application US/09846808
Patent No. US20020064531A1

GENERAL INFORMATION:
APPLICANT: Walker, David H.
TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
TITLE OF INVENTION: Protein Multigene Family
FILE REFERENCE: D6311
CURRENT APPLICATION NUMBER: US/09/846,808
CURRENT FILING DATE: 2001-05-01
PRIOR APPLICATION NUMBER: 60/201,035
PRIOR FILING DATE: 2000-05-01
NUMBER OF SEQ ID NOS: 53
SEQ ID NO 8
LENGTH: 275
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
FEATURE:
OTHER INFORMATION: P28-8 Outer Membrane Protein of
OTHER INFORMATION: Ehrlichia chaffeensis
US-09-846-808-8

Query Match 21.8%; Score 48; DB 10; Length 275;
Best Local Similarity 34.4%; Pred. No. 28;
Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

OY 6 YRYALKSGSVNAPMPENGOTENNMI 37
DB 136 YRYFALARNPSSPTSNNTYVMRNDGVITS 167

RESULT 12
US-10-059-964-28
Sequence 28, Application US/10059964
Patent No. US20020120115A1

GENERAL INFORMATION:
APPLICANT: Rikihisa, Yasuko
APPLICANT: Ohashi, No. US20020120115A1
TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
TITLE OF INVENTION: Chaffeensis
FILE REFERENCE: 22727/04021
CURRENT APPLICATION NUMBER: US/10/059,964
CURRENT FILING DATE: 2002-01-28
EARLIER APPLICATION NUMBER: 09/314,701
EARLIER FILING DATE: 1999-05-19
NUMBER OF SEQ ID NOS: 66
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 28
LENGTH: 275
TYPE: PRT
ORGANISM: Ehrlichia chaffeensis
US-10-059-964-28

Query Match 21.8%; Score 48; DB 12; Length 275;
Best Local Similarity 34.4%; Pred. No. 28;
Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

OY 6 YRYALKSGSVNAPMPENGOTENNMI 37
DB 136 YRYFALARNPSSPTSNNTYVMRNDGVITS 167

RESULT 13
US-09-740-026A-4
Sequence 4, Application US/09740026A
Patent No. US20020081678A1

```

: SOFTWARE:FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 3084
: TYPE: PRT
: ORGANISM: Mus Musculus
: PUBLICATION INFORMATION:
: DATABASE ACCESSION NUMBER: Swissprot P19137
: DATABASE ENTRY DATE: 1990-11-01
: US-09-938-275-4

Query Match          21.8%  Score 48;  DB 10;  Length 3084;
Best Local Similarity 30.0%;  Pred. No. 5e+02;
Matches 15;  Conservative 3;  Mismatches 18;  Indels 14;  Gaps 2;

QY      5  NYRYVALKSGS-----VNAPMPENG-----QTENNIDILMSQEQ 40
      11  1  11  ::::11111  1  1  1
Db      1260  NYEPVOLIKGRARKHVIYMDAPAFENGVRODYEQMKKEEFKYNYSVE 1309

Search completed: December 10, 2002, 10:58:06
Job time : 5.2 secs

```

```

; RESULT 14
; US-10-108-605-71
; Sequence 71, Application US/10108605
; Patent No. US20020160934A1
; GENERAL INFORMATION:
; APPLICANT: Brodus, Julie
; APPLICANT: Stam, Lynn
; APPLICANT: Bachmann, Jane
; APPLICANT: Kamdar, Kim
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCODE
; FILE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
; FILE REFERENCE: 31133B
; CURRENT APPLICATION NUMBER: US/10/108,605
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 03/761,142
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/176,418
; PRIOR FILING DATE: 2000-01-14
; NUMBER OF SEQ ID NOS: 361
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 71
; LENGTH: 1207
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-10-108-605-71

Query Match          21.8%; Score 48; DB 9; Length 1207;
Best Local Similarity 32.4%; Pred. No. 1.6e+02;
Matches 12; Conservative 8; Mismatches 11; Indels 6; Gaps 1;

QY      1 LYKKNRYRYALKSGSVNAPMPENGTENNNDWTLMGS 37
Db       441 LVHRRKKRY-----GHLPALMSNNGPTASNTYQLLNS 471
||: || | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | | |

RESULT 15
; US-09-938-275-4
; Sequence 4, Application US/09938275
; Patent No. US2002011309A1
; GENERAL INFORMATION:
; APPLICANT: Gerardo Castillio
; APPLICANT: Alan Snow
; TITLE OF INVENTION: Therapeutic and Diagnostic Applications
; TITLE OF INVENTION: of Laminin and Laminin-Derived Protein Fragments
; FILE REFERENCE: PROTEO.P03
; CURRENT APPLICATION NUMBER: US/09/938,275
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 11

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:54:49 ; Search time 9.8 Seconds
(Without alignments)
268,490 Million cell updates/sec

Title: US-09-142-970-1_COPY_1_40
Perfect score: 220
Sequence: 1 LYKKNRYALKSGSVNAPMPENGQTENNWMIGSTOB 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 232864 seqs, 65780094 residues

Total number of hits satisfying chosen parameters: 232864

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New.*
1: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep.*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep.*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49.5	22.5	400	1 PCT-US02-32727-3449	Sequence 3449, Ap
2	49.5	22.5	400	6 US-10-057-498-3449	Sequence 3449, Ap
3	48.5	22.0	378	6 US-10-216-209-87	Sequence 87, Appl
4	48	21.8	275	6 US-10-284-986-8	Sequence 8, Appl
5	48	21.8	275	6 US-10-285-042-8	Sequence 8, Appl
6	47.5	21.6	269	5 US-09-724-676-66201	Sequence 66201, A
7	47.5	21.6	269	5 US-09-724-676-66220	Sequence 66220, A
8	47.5	21.6	269	5 US-09-724-676A-66201	Sequence 66201, A
9	47.5	21.6	269	5 US-09-724-676A-66220	Sequence 66220, A
10	47.5	21.6	295	5 US-09-724-676-66201	Sequence 66201, A
11	47.5	21.6	295	5 US-09-724-676-66219	Sequence 66219, A
12	47.5	21.6	295	5 US-09-724-676A-66200	Sequence 66200, A
13	47.5	21.6	295	5 US-09-724-676A-66219	Sequence 66219, A
14	47.5	21.6	348	6 US-10-180-932-2	Sequence 2, Appl
15	47.5	21.6	353	5 US-09-724-676-66198	Sequence 66198, A
16	47.5	21.6	353	5 US-09-724-676-66199	Sequence 66199, A
17	47.5	21.6	353	5 US-09-724-676-66217	Sequence 66217, A
18	47.5	21.6	353	5 US-09-724-676-66218	Sequence 66218, A
19	47.5	21.6	353	5 US-09-724-676A-66198	Sequence 66198, A
20	47.5	21.6	353	5 US-09-724-676A-66199	Sequence 66199, A
21	47.5	21.6	353	5 US-09-724-676A-66217	Sequence 66217, A
22	47.5	21.6	353	5 US-09-724-676A-66218	Sequence 66218, A
23	47	21.4	1075	5 US-09-134-000C-4599	Sequence 4599, Ap
24	47	21.4	1075	5 US-09-724-676-67169	Sequence 67169, A
25	47	21.4	1075	5 US-09-724-676-67170	Sequence 67170, A
26	47	21.4	1075	5 US-09-724-676-67171	Sequence 67171, A

ALIGNMENTS

```
RESULT 1
PCT-US02-32727-3449
; Sequence 3449, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skelky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yann
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 3449
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Propionibacterium
PCT-US02-32727-3449

Query Match
Best local Similarity 22.5%; Score 49.5; DB 1; Length 400;
Matches 13; Conservative 11; Mismatches 14; Indels 23; Gaps 2;

QY 2 YVKNRYALKSGS-----SVNAPMPENGQT--ENNWMIGST 38
DB 250 YVDFVHVNFPDGPYLIHIDATFVPLRGLITNNPRLPEQGRKIFRANQWQVDA 309
QY 39 Q 39
DB 310 Q 310

RESULT 2
US-10-057-498-3449
; Sequence 3449, Application US/10057498
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skelky, Yasir
; APPLICANT: Persing, David
```

```

; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes
; FILE REFERENCE: 210121.514
; CURRENT APPLICATION NUMBER: US/10/057,498
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 3449
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Propionibacterium acnes
US-10-057-498-3449

Query Match
Best Local Similarity 22.5%; Score 49.5; DB 6; Length 400;
Matches 13; Conservative 11; Mismatches 14; Indels 23; Gaps 2;

QY 2 YKYNRYVALKSG-----SYNAPMPENGQT--ENNMDILMGST 38
DB 250 YYPDEFVHAVNPPGDPPIHIDATEFVPLRPGLIINNPNRPLPQEQRKIFEANDQIVDAA 309
QY 39 Q 39
DB 310 Q 310

RESULT 3
US-10-216-209-87
; Sequence 87, Application US/10216209
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: De Kievit, Teresa De
; TITLE OF INVENTION: Novel Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-167
; CURRENT APPLICATION NUMBER: US/10/216,209
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: US/09/352,994
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 08/846,762
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Salmonella enterica
US-10-216-209-87

Query Match
Best Local Similarity 22.0%; Score 48.5; DB 6; Length 378;
Matches 8; Conservative 11; Mismatches 8; Indels 3; Gaps 1;

QY 2 YKYNRYVALKSG---GSYNAPMPENGQTE 28
DB 108 FYONIKVWHLEAGLRWNNMNSPPEEGNRQ 137

RESULT 4
US-10-284-986-8
; Sequence 8, Application US/10284986
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D1
; CURRENT APPLICATION NUMBER: US/10/284,986
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 8
```

```

; LENGTH: 275
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-8 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-284-986-8

Query Match
Best Local Similarity 21.8%; Score 48; DB 6; Length 275;
Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 6 YRYVALKSGSVNAPMPENGQTEENNMDILMG 37
DB 136 YRYFALARNPSSGSSPTSNNTVYMRNDGVSTIS 167

RESULT 5
US-10-285-042-8
; Sequence 8, Application US/10285042
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xu-Jie
; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
; FILE REFERENCE: D6311D2
; CURRENT APPLICATION NUMBER: US/10/285,042
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 09/846,808
; PRIOR FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 53
; SEQ ID NO 8
; LENGTH: 275
; TYPE: PRT
; ORGANISM: Ehrlichia chaffeensis
; FEATURE:
; OTHER INFORMATION: P28-8 Outer Membrane Protein of
; OTHER INFORMATION: Ehrlichia chaffeensis
US-10-285-042-8

Query Match
Best Local Similarity 21.8%; Score 48; DB 6; Length 275;
Matches 11; Conservative 5; Mismatches 16; Indels 0; Gaps 0;

QY 6 YRYVALKSGSVNAPMPENGQTEENNMDILMG 37
DB 136 YRYFALARNPSSGSSPTSNNTVYMRNDGVSTIS 167

RESULT 6
US-09-724-676-66201
; Sequence 66201, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 66201
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676-66201

Query Match
Best Local Similarity 21.6%; Score 47.5; DB 5; Length 269;
Matches 12; Conservative 6; Mismatches 9; Indels 3; Gaps 1;
```

```

QY      11 LKSGGSVNAMP--ENGQTEINDWILMGs 37
          | | | : : | | | : : | | |
Db      16 LPSSGAVDCGIPESIENGKVEDPESTLFGs 45

```

RESULT 7
US-09-724-676-66220
; Sequence 66220, Application US/09724676
; GENERAL INFORMATION.

OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676-66220

Query Match	21.6%	Score 47.5	DB 5	Length 269
Best Local Similarity	40.0%	Pred. No. 73		
Matches 12	Conservative	6	Mismatches 9	Indels 3
				Gaps 1
Qy	11	LKSGGSVNAEMP---ENGCTENNDRILKMS	37	
	16	LPSSGAVDQGIPESLIENGKVEDPESTLFSS	45	

OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676A-66201

```

RESULT 9
US-09-724-676A-66220
: Sequence 66520, Application US/09724676A
: GENERAL INFORMATION:
: APPLICANT: Comphen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comphen

```

: CURRENT APPLICATION NUMBER: US/09/724,676A
 : CURRENT FILING DATE: 2000-11-28
 : NUMBER OF SEQ ID NOS: 97222
 : SOFTWARE: PatentIn version 3.2
 : SEQ ID NO 66220

Query Match	21.68;	Score	47.5;	DB	5;	Length	269;
Best Local Similarity	40.08;	Pred.	No.	73;			
Matches	12;	Conservative	6;	Mismatches	9;	Indels	3;
						Gaps	1.

RESULT 10
US-09-724-676-66200
; sequence 66200, Application US/09724676

Query March	21.6%	Score	47.5	DB	5	Length	295
Best Local Similarity	40.0%	Pred. No.	80				
Matches	12	Conservative	6	Mismatches	9	Indels	3
						Gaps	1

RESULT 11
US-09-724-676-66219
; Sequence 66219, Application US/09724676

OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676-66219

```

Query Match      21.6%; Score 47.5; DB 5; Length 295;
Best Local Similarity 40.0%; Pred. No. 80;
Matches 12; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

OY      11 LKSGSVNAPMP---ENGOTENNDMIILNGS 37
      | | | | | : | | | | | : | | |
Db      16 LPSSGAVDCGIPESIEKGKVEDPEESTLFGS 45

RESULT 12
US-09-724-676A-66200
; Sequence 66200, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 66200
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676A-66200

Query Match      21.6%; Score 47.5; DB 5; Length 295;
Best Local Similarity 40.0%; Pred. No. 80;
Matches 12; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

OY      11 LKSGSVNAPMP---ENGOTENNDMIILNGS 37
      | | | | | : | | | | | : | | |
Db      16 LPSSGAVDCGIPESIEKGKVEDPEESTLFGS 45

RESULT 13
US-09-724-676A-66219
; Sequence 66219, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 66219
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (3)..(3)
; OTHER INFORMATION: Xaa can be any naturally occurring amino acid
US-09-724-676A-66219

Query Match      21.6%; Score 47.5; DB 5; Length 295;
Best Local Similarity 40.0%; Pred. No. 80;
Matches 12; Conservative 6; Mismatches 9; Indels 3; Gaps 1;

OY      11 LKSGSVNAPMP---ENGOTENNDMIILNGS 37
      | | | | | : | | | | | : | | |
Db      16 LPSSGAVDCGIPESIEKGKVEDPEESTLFGS 45

RESULT 14
US-10-180-932-2
; Sequence 2, Application US/10180932

```

```

: GENERAL INFORMATION:
: APPLICANT: Allen, Keith D.
: TITLE OF INVENTION: EDC4 DISRUPTIONS, COMPOSITIONS AND
: TITLE OF INVENTION: METHODS RELATING THERETO
: FILE REFERENCE: R-926
: CURRENT APPLICATION NUMBER: US/10/180.932
: CURRENT FILING DATE: 2002-06-25
: PRIOR APPLICATION NUMBER: US 60/300,979
: PRIOR FILING DATE: 2001-06-26
: NUMBER OF SEQ. ID NOS: 4
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 2
: LENGTH: 348
: TYPE: PR
: ORGANISM: Mus musculus
: US-180-932-2

Query Match          21.6%; Score 47.5; DB 6; Length 348;
Best Local Similarity 37.8%; Pred. No. 95;
Matches 14; Conservative 6; Mismatches 10; Indels 7; Gaps 2.

OY      3 YKNRYRYA-LKSGSVNAPPENGQTNNDWILMST 38
      :| | | | | | | | | | | | | | | |
Db      317 HKSARYSASAOGTGASTRIMLPENGRP-----LMDST 347

RESULT 15
US-09-724-676-66198
: Sequence 66198, Application US/09724676
: GENERAL INFORMATION:
: APPLICANT: Compugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Compugen
: CURRENT APPLICATION NUMBER: US/09/724.676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ. ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ ID NO 66198
: LENGTH: 353
: TYPE: PR
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (3)..(3)
: OTHER INFORMATION: Xaa can be any naturally occurring amino acid
: US-09-724-676-66198

Query Match          21.6%; Score 47.5; DB 5; Length 353;
Best Local Similarity 40.0%; Pred. No. 97;
Matches 12; Conservative 6; Mismatches 9; Indels 3; Gaps 1.

OY      11 LKSGSVNAPMP--ENGQTNNDWILMKS 37
      | | | | | | | | | | | | | | | |
Db      16 LPSSGAVDCGIPESIEENGKVEDPESTLFGS 45

Search completed: December 10, 2002, 11:10:15
Job time : 10 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:50:48 ; Search time 11 Seconds
(without alignments)
349.580 Million cell updates/sec

Title: US-09-142-970-1_COPY_1_40

Perfect score: 220
Sequence: 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220	100.0	1561	2 S61314	IgA-specific metal
2	220	100.0	1773	2 A81937	IgA-specific metal
3	216	98.2	1815	2 C81169	IgA-specific metal
4	190	86.4	1532	2 A26039	IgA-specific metal
5	98	44.5	1541	2 A37023	IgA-specific metal
6	98	44.5	1694	2 H64106	IgA-specific metal
7	98	44.5	1702	2 A41859	IgA-specific metal
8	89	40.5	1849	2 C41859	IgA-specific metal
9	86	39.1	1345	2 B41859	IgA-specific metal
10	59	26.8	359	2 D90603	IgA-specific metal
11	59	26.8	359	2 A43532	IgA-specific metal
12	58.5	25.9	114	2 AD0785	IgA-specific metal
13	57	25.9	452	2 A10517	IgA-specific metal
14	56.5	25.7	291	2 S03230	IgA-specific metal
15	56	25.5	452	2 B90640	IgA-specific metal
16	56	25.5	452	2 B85491	IgA-specific metal
17	56	25.5	452	2 F64730	IgA-specific metal
18	56	25.5	478	2 F90497	IgA-specific metal
19	54.5	24.8	532	1 CPBYX	IgA-specific metal
20	54.5	24.8	600	2 B96990	IgA-specific metal
21	54.5	24.8	600	2 A38437	IgA-specific metal
22	54.5	24.8	600	2 A38437	IgA-specific metal
23	53.5	24.3	293	2 C75421	IgA-specific metal
24	53.5	24.3	1317	2 F83310	IgA-specific metal
25	53	24.1	152	2 H64749	IgA-specific metal
26	53	24.1	159	2 B90979	IgA-specific metal
27	53	24.1	161	2 B90804	IgA-specific metal
28	53	24.1	161	2 G85663	IgA-specific metal
29	53	24.1	300	2 I40425	IgA-specific metal

30	53	24.1	300	2 C69664	carboxylesterase N
31	53	24.1	596	2 T23193	hypothetical prote
32	53	24.1	646	2 T28868	hypothetical prote
33	53	24.1	658	2 E95111	endo-beta-N-acetyl
34	53	24.1	721	2 C97980	endo-beta-N-acetyl
35	52.5	23.9	186	2 G82439	MutR/nudix family
36	52.5	23.9	850	2 S20462	RNAI2 protein - Ye
37	52	23.6	343	2 G96533	replicative DNA he
38	52	23.6	471	2 H82814	glucan 1,6-alpha-l
39	52	23.6	641	2 A55549	endo-1,4-beta-xyla
40	52	23.6	954	2 S20907	hypothetical prote
41	52	23.6	1356	2 T16718	heat shock sigma f
42	51	23.2	298	2 H87632	trehalose-phosphat
43	51	23.2	374	2 T52058	protein hypothetic
44	51	23.2	374	2 H96809	hypothetical prote
45	51	23.2	835	2 C86444	hypothetical prote

ALIGNMENTS

RESULT 1
S61314
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: H713
C>Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61314
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A:Reference number: S61314; PMID:95302961; PMID:7783620
A:Accession: S61314
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1561 <Lom>
A:Cross-references: EMBL:X82474; NID:g732873; PIDN:CA57857.1; PID:g732874
C:Superfamily: IgA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 100.0%; Score 220; DB 2; Length 1561;
Best Local Similarity 100.0%; Pred. No. 9e-21;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
DB 584 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 623

copy to core file

RESULT 2
A81937
IgA-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - Neisseria menin
N:Alternate names: IgA1 protease; IgA1 protease (EC 3.4.21.7) [misnomer]; immunog1
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain H717; strain H719; strain SM1027
C>Date: 05-May-2000 #sequence-revision 05-May-2000 #text-change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
; Holroyd, S.; Jagsels, K.; Leather, S.; Moulle, S.; Mungall, K.; Quail, M.A.; Rajandre
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; PMID:20222556; PMID:10761919
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737
A:Experimental source: serogroup A, strain 22491
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IgA1 protease in Neiss
A:Reference number: S61314; PMID:95302961; PMID:7783620
A:Accession: S61317
A:Status: nucleic acid sequence not shown

A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82470; NID:g732854; PIDN:CAA57853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82471; NID:g732856; PIDN:CAA57854.1; PID:g732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82472; NID:g732852; PIDN:CAA57855.1; PID:g732853
A:Experimental source: strain SM1027
A:Gene: iga; NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 100.0%; Score 220; DB 2; Length 1773;
Best Local Similarity 100.0%; Pred. No. 1e-20;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
DB 573 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 612

RESULT 3

Iga-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [Imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: strain MC58 serogroup B; strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Retellin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Haft, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <TEXT>
A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AAF41117.1; PID:g722592
A:Experimental source: serogroup B, strain MC58
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseri*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82477; NID:g732856; PIDN:CAA57860.1; PID:g732857
C:Genetics:
A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 98.2%; Score 216; DB 2; Length 1815;
Best Local Similarity 97.5%; Pred. No. 3.6e-20;
Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
DB 573 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 612

RESULT 4

A26039
Iga-specific metalloendopeptidase (EC 3.4.24.13) precursor - *Neisseria gonorrhoeae* (s
N:Alternate names: Iga protease; immunoglobulin A1 proteinase
C:Species: *Neisseria gonorrhoeae*
A:Variety: strain MS11
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000
R:Polnert, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 458-462, 1987
A:Title: Gene structure and extracellular secretion of *Neisseria gonorrhoeae* Iga prot
A:Reference number: A26039; MUID:87115823; PMID:3027577
A:Accession: A26039
A:Molecule type: DNA
A:Residues: 1-1532 <POH>
A:Cross-references: GB:X04835; NID:g44868; PIDN:CAA26538.1; PID:g44869
A>Note: the authors translated the codon AAG for residue 668 as Asn
R:Halter, R.; Polnert, J.; Meyer, T.F.
EMBO J. 8, 2737-2744, 1989
A:Title: Mosaic-like organization of Iga protease genes in *Neisseria gonorrhoeae* gene
A:Reference number: S09386; MUID:9060036; PMID:2511009
A:Accession: S09386
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'W', 429-531, 'N', 533-615, 'V', 617-631, 'N', 6
A:Experimental source: strain MS11
C:Genetics:
A:Gene: iga
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>
F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match 86.4%; Score 190; DB 2; Length 1532;
Best Local Similarity 85.0%; Pred. No. 8.7e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
DB 584 LYKKNRYRYALKSGRLNAPMPENGVAENNDWIFMGVTOE 623

RESULT 5

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - *Haemophilus influ*
N:Alternate names: immunoglobulin A1 proteinase type 1
C:Species: *Haemophilus influenzae*
C:Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Killian, M.
Infect. Immun. 57, 3097-3105, 1989
A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of *Haemo*
A:Reference number: A37023; MUID:89379374; PMID:2506130
A:Accession: A37023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1541 <POH>
A:Cross-references: GB:X64357; NID:g43560; PIDN:CAA45708.1; PID:g43561
A:Experimental source: serotype b
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 44.5%; Score 98; DB 2; Length 1541;
Best Local Similarity 40.0%; Pred. No. 0.00015;
Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYKKNRYRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
DB 595 LNLNRYRYALKGASTRSELPKNSGSESNWLYMGKTSO 634

RESULT 6
H64106

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain N:Alternate names: immunoglobulin A1 proteinase type 1

A:Variety: strain Rd KW20

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000

C:Accession: H64106; A41500

R:Fieldmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodde, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64106

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1694 <TIGR>

A:Cross-references: GB:032779; GB:I42023; NID:g1574009; PIDN:AAC22651.1; PID:g1574019; T

R:Grundy, F.J.; Plant, A.G.; Wright, A.

Infect. Immun. 58, 320-331, 1990

A:Title: Localization of the cleavage site specificity determinant of Haemophilus influenzae Rd.

A:Reference number: A41500; MUID:90129281; PMID:2105270

A:Accession: A41500

A:Molecule type: DNA

A:Residues: 1-377 <GRU>

A:Cross-references: GB:X59800

A:Experimental source: strain Rd KW20

A:Note: the authors translated the codon TGG for residue 319 as Thr

C:Function: A:Description: this proteinase is classified as type 1 because it cleaves at a proline-S

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 44.5%; Score 98; DB 2; Length 1694;

Best Local Similarity 40.0%; Pred. No. 0.00017;

Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGSVNAPMPENGOTENNMDILMGSTOE 40

DB 601 LNFENYTYALRKAGSTRSELPKNGSGENWLYMGKTS 640

RESULT 7
A41859

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae

C:Species: Haemophilus influenzae

A:Variety: strain HK715

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: A41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: A41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1702 <POU>

A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907

A:Experimental source: strain HK715

A:Note: sequence extracted from NCBI backbone (NCBIP:97282)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 44.5%; Score 98; DB 2; Length 1702;

Best Local Similarity 40.0%; Pred. No. 0.00017;

Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGSVNAPMPENGOTENNMDILMGSTOE 40

DB 601 LNFENYTYALRKAGSTRSELPKNGSGENWLYMGKTS 640

RESULT 8
C41859

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

A:Variety: strain HK613

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: C41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: C41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613

A:Note: sequence extracted from NCBI backbone (NCBIP:97285)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 40.5%; Score 89; DB 2; Length 1849;

Best Local Similarity 38.1%; Pred. No. 0.003;

Matches 16; Conservative 7; Mismatches 17; Indels 2; Gaps 1;

OY 1 LYYK--NRYVALKSGSVNAPMPENGOTENNMDILMGSTOE 40

DB 604 LYFODNRSYTYTLKKGASTRSELPONSGSENWLYMGKTS 645

RESULT 9
B41859

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influ

C:Species: Haemophilus influenzae

A:Variety: strain HK393

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: B41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: B41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1545 <POU>

A:Cross-references: GB:M87490; NID:g148908; PIDN:AAA24967.1; PID:g148909

A:Experimental source: strain HK393

A:Note: sequence extracted from NCBI backbone (NCBIP:97283)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 39.1%; Score 86; DB 2; Length 1545;

Best Local Similarity 41.0%; Pred. No. 0.0061;

Matches 16; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

OY 1 LYY--KNRYVALKSGSVNAPMPENGOTENNMDILMG 37

DB 597 LYFNEENTYALKKKASINSEFPONSGSENWLYMG 635

RESULT 10
D90603

hypothetical protein MYPU_7320 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C:Accession: D90603

R:Chamand, I.; Heilig, R.; Ferris, S.; Barbe, V.; Samson, D.; Gallison, F.; Moszer,

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: D90603

A:Status: preliminary

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gend
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90640
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833513.1; PID:g13359546; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs0090

Query Match	25.5%;	Score 56;	DB 2;	Length 452;
Best Local Similarity	38.5%;	Pred. No. 15;		
Matches 10;	Conservative 6;	Mismatches 6;	Indels 4;	Gaps 1;

.Qy	12	KSGGSYNAPMPENG----	OTENNNDWI	33
Db	200	KAKGEIFGILPENGIAIMNADNDWL	225	

Search completed: December 10, 2002, 10:57:35
Job time : 14 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:17:38 ; Search time 5.6 seconds

(without alignments)
296.259 Million cell updates/sec

Title: US-09-142-970-1_COPY_1_40

Sequence: 1 LYKKNRYRYALKSGSVNAPMPENCOTENNIMLGSTOE 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	86.4	1532	1	IGA_NEIGO
2	98	44.5	1541	1	IGA1_HAEIN
3	98	44.5	1694	1	IGA0_HAEIN
4	98	44.5	1702	1	IGA2_HAEIN
5	89	40.5	1849	1	IGA4_HAEIN
6	86	39.1	1545	1	IGA3_HAEIN
7	59	26.8	359	1	CD72_HUMAN
8	56.5	25.7	291	1	Y3ZK_SSV1
9	56	25.5	452	1	MORF_ECOLI
10	54.5	24.8	532	1	CBPY_YEAST
11	54.5	24.8	874	1	SVY_STRCO
12	54.5	24.8	993	1	TSH_STRCO
13	54	24.5	922	1	GUND_CLOCL
14	53	24.1	152	1	YARX_ECOLI
15	53	24.1	658	1	LYTB_STRPN
16	52.5	23.9	850	1	RNI2_YEAST
17	52	23.6	641	1	IMD_ARTGO
18	52	23.6	954	1	XYNA_RUMEL
19	51.5	23.4	510	1	CEST_PIG
20	51	23.2	295	1	RP32_CAUCR
21	51	23.2	1319	1	BCC1_ACEXY
22	50.5	23.0	114	1	YEOG_ECOLI
23	50.5	23.0	394	1	LIP3_DROME
24	50.5	23.0	593	1	CGP1_CAEEL
25	50	22.7	230	1	CLD2_CANPA
26	50	22.7	342	1	RLPA_CSEAB
27	50	22.7	1325	1	BCC3_ACEXY
28	49.5	22.5	507	1	YME3_YEAST
29	49	22.3	269	1	ESL1_MYCPN
30	49	22.3	297	1	YMY9_YEAST
31	49	22.3	382	1	ACP2_ZYMMO
32	49	22.3	389	1	ACDP_MICLE
33	49	22.3	512	1	SING_DROME

34	49	22.3	575	1	CYD2_BACSU	P94367 bacillus su
35	49	22.3	638	1	GYR8_TREDE	O87545 treponema d
36	49	22.3	1660	1	VITE_OSCBR	O94637 oscineus br
37	48.5	22.0	253	1	RSO_CANAL	O42817 candida alb
38	48.5	22.0	378	1	REBC_SALBO	P52642 salmonella
39	48.5	22.0	448	1	SPS2_HUMAN	O99611 homo sapien
40	48.5	22.0	452	1	SPS2_MOUSE	P97364 mus musculu
41	48	21.8	66	1	TXMA_DENPO	P80494 dendroaspis
42	48	21.8	139	1	RBS_CYLSN	P24683 cyllindrothe
43	48	21.8	139	1	RBS_THANO	O98948 thalassiosl
44	48	21.8	318	1	IKBA_CHICK	O91974 gallus gall
45	48	21.8	327	1	ACCO_DORSP	P31238 dorilaenops

ALIGNMENTS

RESULT 1	IGA_NEIGO	STANDARD:	PRT:	1532 AA.
DT	01-MAR-1989 (Rel. 10, Created)			
DT	01-MAR-1989 (Rel. 10, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA protease).			
OS	Neisseria gonorrhoeae.			
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.			
NC	NCBI_TaxID=485;			
LN	(1)			
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.			
RC	STRAIN=MS11.			
RA	MEDLINE=87115823; PubMed=3027577;			
RT	Pohlner J., Hatter R., Beyreuther K., Meyer T.F.;			
RT	"Gene structure and extracellular secretion of Neisseria gonorrhoeae Iga protease.";			
RL	Nature 325:458-462(1987).			
RP	ACTIVE SITE.			
RA	MEDLINE=90154052; PubMed=2105953;			
RT	Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;			
RT	"Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and Hemophilus influenzae by peptide prolyl boronic acids.";			
RL	J. Biol. Chem. 265:3738-3743(1990).			
CC	-1- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.			
CC	-1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at			
CC	certain Pro-I-Xaa bonds in the hinge region. No small molecule			
CC	substrates are known.			
CC	-1- SUBCELLULAR LOCATION: Secreted.			
CC	-1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC			
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE			
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE			
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.			
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb.ch/announce/			
CC	or send an email to license@isb.stb.ch).			
CC	-----			
DR	EMBL: X04835; CA28538.1; -			
DR	PIR: A26039; A26039.			
DR	MEROPS: S06.001; -			
DR	InterPro: IPR000710; IGA_S6.			
DR	InterPro: IPR004899; Peptact_sup.			
DR	Pfam: PF02395; IGA1; 1.			
DR	Pfam: PF03212; Peptactin; 1.			
DR	PRINTS: PR00921; IGASERPRASE.			
KW	Hydrolase; Serine protease; zymogen; Autocatalytic cleavage;			

KW	Transmembrane; Signal.	
FT	SIGNAL	1 27
FT	CHAIN	28 986
FT	PROPEP	987 1532
FT	ACT_SITE	278 278
FT	SITE	986 987
FT	SITE	1018 1019
FT	SITE	1121 1122
SO	SEQUENCE	1532 AA; 168976 MW; 68FF4112BD2FF40D CRC64;
Query Match		
Best Local Similarity	86.4%;	Score 190; DB 1; Length 1532;
Matches	34; Conservative	1; Mismatches 5; Indels 0; Gaps 0;
OY	1	LYKKNYRYALKSGSVNAPMPENQOTENDMDITLGSFTQE 40
DB	584	LYKKNYRYALKSGGRLLNAPMPENQVAENDDITFMGTQE 623

ID	IGAL_HAEIN	STANDARD:	PRT:	1541 AA.
AC	P42782;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL1 protease).			
GN	IGA.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision: Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBI_TaxID=727;			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HK368 / Serotype B;			
RX	MEDLINE=89379374; PubMed=2506130;			
RA	Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;			
RT	"Cloning and sequencing of the immunoglobulin A1 protease gene (iga)			
RL	of Haemophilus influenzae serotype b.;"			
RL	Infect. Immun. 57:3097-3105(1989).			
RN	(2)			
RP	MUTAGENESIS OF SER-288.			
RC	STRAIN=HK368 / Serotype B;			
RX	MEDLINE=92234949; PubMed=1373717;			
RA	Poulsen K., Reimhold J., Kilian M.;			
RT	"A comparative genetic study of serologically distinct Haemophilus			
RL	influenzae type 1 immunoglobulin A1 proteases.;"			
RL	J. Bacteriol. 174:2913-2921(1992).			
CC	-I- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A			
CC	PRODUCING INTACT FC AND F&B FRAGMENTS.			
CC	-I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at			
CC	certain Pro-1-Xaa bonds in the hinge region. No small molecule			
CC	substrates are known.			
CC	-I- SUBCELLULAR LOCATION: Secreted.			
CC	-I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC			
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE			
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE			
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY			
CC	SIMILARITY).			
CC	-I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/).			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X64357; CAA45708.1; -			
DR	EMBL; M87492; AAA24969.1; -			
DR	MEROPS; S06.001; -			
DR	InterPro; IPR000710; IGA_S6.			

```

DR   InterPrct      IP0004899; Pectactl_sup.
DR   Pfam: PF02395; IGA1; 1.
DR   Pfam: PF03212; Pectactin; 2.
DR   PRINTS; PR00921; IGASEPTASE.
KW   Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT   SIGNAL        1
FT   CHAIN         26  1008  POTENTIAL.
FT   PROPEP        1009  1541  IMMUNOGLOBULIN A1 PROPEASE.
FT   ACT SITE      288      288  HELPER PEPTIDE (POTENTIAL).
FT   MUTAGEN       288      288  PROBABLE.
FT   MUTAGEN       288      288  S->T: LOSS OF ACTIVITY.
SQ   SEQUENCE      1541 AA; 169370 MW; CE7257CB3196C600 CAC64;

Query Match          44.5%; Score 98; DB 1; Length 1541;
Best Local Similarity 40.0%; Pred. No. 7,4e-05;
Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

QY   1 LYYKNRYVALKSGSVNAPMENDQGTENNMMIIGSGQE 40
      |::|::|::|::|::|::|::|::|::|::|::|::|
DB   595 LNEENTYYIALRKGASTRSELPEKNGSENNMLYMGKTD 634

```

RESULT 3			
IGAO_HAEIN	STANDARD:	PRT:	1694 AA.
ID	IGAO_HAEIN		
AC	P44969;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAI protease).		
GN	IGA OR IGAI OR HI0990.		
OS	Haemophilus influenzae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;		
OC	Haemophilus.		
OX	NCBI_TaxID=727;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Serotype D;		
RA	Wright A, Fishman Y., Tai F., Plant A.G.;		
RL	Submitted (MAY-1991) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-Rd / KW20 / ATCC 51907;		
RX	MEDLINE=95350630; PubMed=7542800;		
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirsch E.F.,		
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,		
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,		
RA	Scott J.D., Shirley R., Liu L.-I., Glodex A., Kelley J.M.,		
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,		
RA	Ulfenback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,		
RA	Fine L.D., Fritchman J.L., Fuhrmann J.L., Georgagen N.S.M.,		
RA	Greim C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,		
RA	Venter J.C.;		
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae		
RT	Rd.";		
RL	Science 269:496-512(1995).		
CC	-i- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A		
CC	PRODUCING INTACT FC AND FAB FRAGMENTS.		
CC	-i- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at		
CC	certain Pro-I-Xaa bonds in the hinge region. No small molecule		
CC	substrates are known.		
CC	-i- SUBCELLULAR LOCATION: Secreted.		
CC	-i- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC		
CC	SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE		
CC	OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE		
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY		
CC	SIMILARITY).		
CC	-i- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.		
CC	-----		
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration -		
CC	between the Swiss Institute of Bioinformatics and the EMBL collaboration -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	institutions is not permitted.		


```
CC entries requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X59800; NOT_ANNOTATED_CDS.
DR EMBL: U32779; AAC22651.1; -.
DR MEROPS: S06.001; -.
DR TIGR: H10990; -.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF03212; Pertactin_2.
DR Pfam: PF03212; IGASERPTASE.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 25
FT CHAIN 1 25
FT PROPEP 26 1014
FT ACT_SITE 1015 1694
FT ACT_SITE 288 288
FT CONFLICT 253 254
FT CONFLICT 272 272
FT CONFLICT 464 464
FT CONFLICT 866 866
FT CONFLICT 1036 1036
FT CONFLICT 1074 1074
FT CONFLICT 1421 1421
FT CONFLICT 1545 1545
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 44.5%; Score 98; DB 1; Length 1694;
Best Local Similarity 40.0%; Pred. No. 8.2e-05;
Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGSVNAPMPGOTENNNDWILMGSTOE 40
DB 601 LNEHTYYALRKGASTRSELPKNGSENEWLMYMGKTS 640

RESULT 4
IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT Influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M87489; AAA24966.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF03212; Pertactin_2.
DR Pfam: PF03212; IGASERPTASE.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
KW Complete proteome.
FT SIGNAL 1 25
FT CHAIN 1 25
FT PROPEP 26 1014
FT ACT_SITE 1015 1702
FT ACT_SITE 288 288
FT DOMAIN 1109 1124
FT REPEAT 1109 1116
FT REPEAT 1117 1124
FT REPEAT 1117 1124
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;

Query Match 44.5%; Score 98; DB 1; Length 1702;
Best Local Similarity 40.0%; Pred. No. 8.2e-05;
Matches 16; Conservative 8; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGSVNAPMPGOTENNNDWILMGSTOE 40
DB 601 LNEHTYYALRKGASTRSELPKNGSENEWLMYMGKTS 640
```

```
RESULT 5
IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH71 HK61;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT Influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
```

```

CC or send an email to license@sib.ch).
CC -----
DR EMBL: M87491: AAA24968.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710: IGA_S6.
DR InterPro: IPR004899: Pectact_sup.
DR Pfam: PF032395; IGA1; 1.
DR Pfam: PF03212; Pectactin; 2.
DR PRINTS: PR00921: IGASERPTASE.
KM Hydrolyase: Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1021
FT PROPEP 1022 1849
FT ACT_SITE 299 299
FT ACT_SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 40.5%; Score 89; DB 1; Length 1849;
Best Local Similarity 38.1%; Pred. No. 0.0014;
Matches 16; Conservative 7; Mismatches 17; Indels 2; Gaps 1;

QY 1 LYYK--NYRYVALKSGSVNAPMPENGQTENNMDWILMGSTGE 40
Db 604 LYFMDNRSYITLKKGASTSELPQNGSGENKMLYMGRTSD 645

RESULT 6
IGA1_HAEIN STANDARD: PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
ON NCBI_TaxID=727;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK93 / NCTC 8467 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RA "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC -1- PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M87490: AAA24967.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710: IGA_S6.
DR InterPro: IPR004899: Pectact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pectactin; 2.

```

```

DR PRINTS: PR00921: IGASERPTASE.
KM Hydrolyase: Serine protease; Transmembrane; zymogen; signal.
FT SIGNAL 1 25
FT CHAIN 26 1012
FT PROPEP 1013 1545
FT ACT_SITE 292 292
FT ACT_SITE 292 292
SQ SEQUENCE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;

Query Match 39.1%; Score 86; DB 1; Length 1545;
Best Local Similarity 41.0%; Pred. No. 0.0028;
Matches 16; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 1 LYY--KNRYVALKSGSVNAPMPENGQTENNMDWILMGST 37
Db 597 LYFNEENRTYVALKKDASIRSEFPQNGSGENNSMLYMGST 635

RESULT 7
CD72_HUMAN STANDARD: PRT; 359 AA.
ID CD72_HUMAN
AC P21854;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE B-cell differentiation antigen CD72 (Lyb-2).
GN CD72.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
ON NCBI_TaxID=9606;
RX [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=90278102; PubMed=2141045;
RA von Hoegen I., Nakayama E., Parnes J.R.;
RT "Identification of a human protein homologous to the mouse Lyb-2 B
RT cell differentiation antigen and sequence of the corresponding
RT cDNA.";
RL J. Immunol. 144:4870-4877(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Spleen;
RA Strausberg R.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A ROLE IN B CELL PROLIFERATION AND
CC DIFFERENTIATION. ASSOCIATES WITH CD5.
CC -1- SUBUNIT: HOMODIMER. DISULFIDE-LINKED.
CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
CC -1- TISSUE SPECIFICITY: PRE-B-CELLS AND B-CELLS BUT NOT TERMINALLY
CC DIFFERENTIATED PLASMA CELLS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD72 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd72.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL: M54992: AAA36189.1; -.
DR EMBL: BC030227: AAH30227.1; -.
DR PIR: A43532; A43532.
DR Genew: HGNC:1696; CD72.
DR MIM: 107272; -.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN 1; FALSE_NEG.
DR PROSITE: PS50041; C-TYPE LECTIN 2; 1.
KW Antigen; B-cell; Glycoprotein; Transmembrane; Lectin; Signal-anchor.
FT DOMAIN 1 95
FT CYTOPLASMIC (POTENTIAL).

```

CC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
OX NCBI_taxid=562;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-69345095; PubMed-2668880.
RA Parquet C., Flouret B., Mengin-Lecreulx D., van Heijenoort J.;
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNAc-
tripeptide synthetase of Escherichia coli.";
RL Nucleic Acids Res. 17:5379-5379(1989).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RT Isono K., Moribuchi K., Nakata A.;
RL "Systematic sequencing of the Escherichia coli genome: analysis of
the 0-2.4 min region.";
RN Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN-MB2884;
RX MEDLINE-92128642; PubMed=89732200.
RA Anderson M.S., Eveland S.S., Onishi H.R., Pompilano D.L.;
RT "Kinetic mechanism of the Escherichia coli UDPMurNAc-tripeptide
fusion.";
RL Biochemistry 35:16264-16269(1996).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed=9278503;
RA Blatter F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RN SEQUENCE FROM N.A. AND MUTANT MURF2.
RC STRAIN-CGSC 5990;
RX MEDLINE-97309380; PubMed=916795.
RA Eveland S.S., Pompilano D.L., Anderson M.S.;
RT "Conditionally lethal Escherichia coli murein mutants contain point
defects that map to regions conserved among murein and folyl poly-
gamma-glutamate ligases: identification of a ligase superfamily.";
RL Biochemistry 36:6223-6229(1997).
RN [6]
RN CHARACTERIZATION, AND SEQUENCE OF 1-15.
RX MEDLINE-90248455; PubMed=2186811;
RA Duncan K., van Heijenoort J., Walsh C.T.;
RT "Purification and characterization of the D-alanyl-D-alanine-adding
enzyme from Escherichia coli.";
RL Biochemistry 29:2379-2386(1990).
RN [7]
RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE-20545602; PubMed=11090285;
RA van Y., Munshi S., Leitling B., Anderson M.S., Chrzas J., Chen Z.;
RT "Crystal structure of Escherichia coli UDPMurNAc-tripeptide
D-alanyl-D-alanine-adding enzyme (murF) at 2.3-A resolution.";
RL J. Mol. Biol. 304:435-445(2000).
RN -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYZES THE FINAL
STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
PRECURSOR OF MUREIN.
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
glutamate-meso-2,6-diaminohexanedioate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-alanine = ADP
CC + carboxyl-L-lysyl-D-alanyl-D-alanine.
CC -1- PATHWAY: Peptidoglycan biosynthesis.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE MURDEF FAMILY.

```
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; X15432; CA93473.1; -
DR      EMBL; X55034; CA93863.1; -
DR      EMBL; D10483; BA01351.1; -
DR      EMBL; U67891; AAC4457.1; -
DR      EMBL; AE000118; AAC73197.1; -
DR      EMBL; U67893; AAB60788.1; -
DR      PIR; S04846; S04846.
DR      PIR; S40596; S40596.
DR      PDB; 1GG4; 20-DEC-00.
DR      Ecogen; EG10622; murF.
DR      InterPro; IPR000713; Mur_ligase_C.
DR      InterPro; IPR004101; Mur_ligase_C.
DR      Pfam; PF01225; Mur_ligase_1.
DR      Pfam; PF08275; Mur_ligase_C; 1.
DR      TIGRfams; TIGR01143; murF_1.
KW      Peptidoglycan synthetis; Cell division; Cell wall; Ligase;
KW      ATP-binding; 3D-structure; Complete proteome.
FT      NP_BIND          107    113
FT      VARIANT         288    288
FT      ACTIVITY        A -> T (IN MURP); TS MUTANT WITH LOW
FT      CONFLICT        61     61
FT      CONFLICT        178    178
SQ      SEQUENCE        452 AA; 47447 MW; B46E2B57BDBC572 CRC64;
Query Match              25.5%; Score 56; DB 1; Length 452;
Best Local Similarity   38.5%; Pred. No. 6.4;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;
Oy      12 KSGSVNAPMPNG-----QTENNMT 33
Db      200 KAKGEIFGSLPENGIAIMNADNDML 225
RESULT 10
CBPY_YEAST
ID      CBPY_YEAST      STANDARD:      PRT;      532 AA.
AC      P00729;
DT      21-JUL-1986 (Rel. 01, Created)
DT      01-NOV-1988 (Rel. 09, Last sequence update)
DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YscY).
GN      PCRI OR YMR297W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI TaxID=4932;
LN      [1]
RP      SEQUENCE FROM N.A.
RC      MEDLINE=87131100; PubMed=3028649;
RA      Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;
RT      "Protein sorting in yeast: the localization determinant of yeast
RL      vacuolar carboxypeptidase Y resides in the propeptide.";
RL      Cell 48:887-897(1987).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=S288c / AB972;
RA      Badcock K., Churcher C., Barrel B.G., Rajandream M.A.;
RT      Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE OF 112-532.
RA      Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;
RT      "Amino acid sequence of carboxypeptidase Y. II. Peptides from
FT      enzymatic cleavages.";
FT      Carlsberg Res. Commun. 47:15-27(1982).
LN
```

[4]
 REVISIONS, AND ACTIVE SITE SER-257.
 Bredam K., Svendsen I.;
 "Identification of methionyl and cysteinyl residues in the substrate
 binding site of carboxypeptidase Y.";
 Carlsberg Res. Commun. 49:639-645(1984).
 [5]
 ACTIVE SITE HIS-508.
 MEDLINE=90315013; PubMed=2639680;
 Bech L.M., Bredam K.;
 "Inactivation of carboxypeptidase Y by mutational removal of the
 putative essential histidyl residue.";
 Carlsberg Res. Commun. 54:165-171(1989).
 [6]
 MUTAGENESIS.
 MEDLINE=94114535; PubMed=7904479;
 Mortensen U.H., Remington S.J., Bredam K.;
 "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen
 bond network stabilizes the transition state by interaction with the
 C-terminal carboxylate group of the substrate.";
 Biochemistry 33:508-517(1994).
 [7]
 X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
 MEDLINE=95244421; PubMed=7727362;
 Endrizzi J.A., Bredam K., Remington S.J.;
 "2.8-A structure of yeast serine carboxypeptidase.";
 Biochemistry 33:11106-11120(1994).
 -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS
 PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC
 RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR
 PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
 -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 broad specificity.
 -1- ENZYME REGULATION: INHIBITED BY 2PCK.
 -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
 -1- PM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMAGEN AND
 IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO
 A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE
 GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE
 GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
 SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS
 SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR
 ENDOSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
 N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY
 ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).
 -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S10.
 -1- DATABASE: NAME=Worthington enzyme manual;
 WWW="http://www.worthington-biochem.com/manual/C/COY.html".

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: M15482; AAA34902.1; -
 EMBL: X80836; CAA56806.1; -
 PIR: A26597; CPBYT.
 DR PDB: 1YSC; 22-JUN-94.
 DR PDB: 1CPY; 15-SEP-95.
 DR MEROPS: S10.001; -
 DR SGD: S0004912; PRCL.
 DR InterPro: IPR000379; Ser_estrs_site.
 DR InterPro: IPR001563; Serine_carpept.
 DR Pfam: PF00450; serine_carpept.1.
 DR PRINTS: PR00724; CROXYPTASEC.
 DR ProDom: PD001181; Serine_carpept.1.
 DR PROSITE: PS00133; CARBOXYPEPT_SER_SER; 1.
 DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
 DR HydroLase: Carboxypeptidase; Glycoprotein; Zymogen; Signal;
 3D-structure-

RT	"Hemoeitic complex and leashirt genes co-operate to establish trunk
Rt	segmental identities in Drosophila.";
RL	Development 120:2287-2296(1994)
CC	-I- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF DROSOPHILA EMBRYONIC TRUNK SEGMENTS. REPRESS THE EXPRESSION OF HEAD HOMEOITC GENES.
CC	NECESSARY, IN COMBINATION WITH SCR, FOR THE FORMATION OF THE PROTHORACIC SEGMENT. MAY BE INVOLVED IN CHROMATIN STRUCTURE FOR MODULATION OF TRANSCRIPTION. NEGATIVELY CONTROL THE EXPRESSION OF MOO AND POSITIVELY THAT OF DLL AND OF ITS OWN EXPRESSION.
CC	-I- SUBCELLULAR LOCATION: Nuclear.
CC	-I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE TRUNK (FROM P53 TO P513).
CC	-I- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL AND ADULT DEVELOPMENT. NOT MATERIANALLY EXPRESSED.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - I-
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isdb.ch/announce/ or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M57496; AAA28983.1; -
DR	PIR; A38437; A38437.
DR	TRANSFAC; T00805; -
DR	IYBase; FBgn0003866; tsh.
DR	InterPro; IPR000822; znf_C2H2.
DR	Pfam; PF00096; zf-C2H2; 3.
DR	SMART; SM00355; znf_C2H2; 3.
DR	PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR	PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
KW	Developmental protein; Transcription regulation; Repressor; Activator; Zinc-finger; Metal-binding; Repeat; DNA-binding; Nuclear protein.
FT	DOMAIN 104 136 ALA-RICH.
FT	DOMAIN 175 183 ASP/GLU-RICH (ACIDIC).
FT	DOMAIN 354 557 ZINC FINGERS.
FT	ZN_FING 354 378 C2H2-TYPE.
FT	ZN_FING 466 490 C2H2-TYPE.
FT	ZN_FING 533 557 C2H2-TYPE.
FT	DOMAIN 104 107 POLY-ALA.
FT	DOMAIN 115 122 POLY-ALA.
FT	DOMAIN 175 180 POLY-GLU.
FT	DOMAIN 401 407 POLY-PRO.
FT	DOMAIN 830 834 POLY-ASN.
SQ	SEQUENCE 993 AA; 106206 MW; 2DF9C6774FF6BB6D1 CRC64;
Oy	Query Match 24.8%; Score 54.5; DB 1; Length 993;
	Best Local Similarity 35.3%; Pred. No. 24;
	Matches 12; Conservative 6; Mismatches 7; Indels 9; Gaps 1;
Dy	2 YKKNRYIALKSGS-----VNAPPEPQG 26 : :: :: Db 740 YYGHHRYTSSRSRGSCSAEARPLDAIPPEKQQ 773
RESULT 13	
GUND_CLOCL	
ID_GUND_CLOCL	STANDARD; PRT; 515 AA.
AC	P28623;
DT	01-DEC-1992 (Rel. 24, Created)
DT	01-DEC-1992 (Rel. 24, Last sequence update)
DT	01-JUN-1994 (Rel. 29, Last annotation update)
DE	Endoglucanase D precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase C) (Cellulase D).
GN	ENGd.
OS	Clostridium cellulovorans.
OC	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
OC	Clostridium.
OX	NCB1_TaxID=1493;
RM	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 32-44. STRAIN=ATCC 35296;

```

RA Hamamoto T., Poong F., Shoseyo'00, Doi R.H.;
RA Hamamoto T., Poong F., Shoseyo'00, Doi R.H.;
RT "Analysis of functional domains of endoglucanases from Clostridium
RT cellulovorans by gene cloning, nucleotide sequencing and chimeric
RT protein construction.";
RL Mol. Gen. Genet. 231:472-479(1992).
CC -I- FUNCTION: HAS ENDOGLUCANASE ACTIVITY ON CARBOXYMETHYL-CELLULOSE
CC (CMC), CELLULOSEDISE ACTIVITY ON P-NITROPHENYL-CELLULOSE
CC (p-NPC), AND PARTIAL HYDROLYTIC ACTIVITY ON CRYSTALLINE CELLULOSE
CC (AVICEL).
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose.
CC -I- SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).
CC -I- SIMILARITY: CONTAINS 1 BACTERIAL-TYPE CELLULOSE-BINDING DOMAIN
CC (CBD).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, M37434; AAA23233.1; -
CC DR HSSP; P07986; 1EXG.
CC DR InterPro: IPR001919; Bac_cellose-Bind.
CC DR InterPro: IPR001547; GH_5..
CC DR InterPro: IPR001230; Prenyl-site.
CC DR Pfam; PF00150; cellulase; 1.
CC DR PROSITE; PS00659; GLYCOSYL-HYDROL_F5; 1.
CC KW Cellulose degradation; Hydrolase; Glycosidase; Signal.
CC FT SIGNAL 1 31
CC FT CHAIN 32 515
CC FT DOMAIN 32 376 ENDOGLUCANASE D.
CC FT DOMAIN 377 407 CATALYTIC (BY SIMILARITY).
CC FT DOMAIN 408 515 PRO/THR-RICH (LINKER).
CC FT ACT_SITE 180 180 CELLULOSE-BINDING.
CC FT ACT_SITE 303 303 PROTON DONOR (BY SIMILARITY).
CC FT ACT_SITE 303 303 NUCLEOPHILE (BY SIMILARITY).
CC FT SEQUENCE 515 AA; 55976 MW; 4CEB736CE76373FO CRC64;
CC SQ
CC
CC Query Match 24.5%; Score 54; DB 1; Length 515;
CC Best local Similarity 50.0%; Pred. No. 14;
CC Matches 11; Conservative 1; Mismatches 10; Indels 0; Gaps 0;
CC
CC QY 13 SGGSYNAAPPENGQGTENNNDWIL 34
CC II III : II I I I I
CC DB 423 SGGSYNAVTKKNGTTPPINGWTL 444
CC
CC RESULT 14
CC YAFX_ECOLI STANDARD: PRT; 152 AA.
CC ID YAFX_ECOLI STANDARD:
CC AC P75676; P11286; Q9R2D9:
CC DT 01-NOV-1997 (Rel. 35, Created)
CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Hypothetical protein yafx.
CC GN YAFX OR B0248.
CC OS Escherichia coli.
CC OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
CC Escherichia.
CC OX NCBI_TaxID=562;
CC [1]
CC RM SEQUENCE FROM N.A.
CC RC STRAIN=K12 / MG1655;
CC RX MEDLINE=97426617; PubMed=9278503;
CC RX Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
CC Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
CC Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
CC Mau B., Shao Y.;

```

```
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / W3110:
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.,
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (189,987 - 281,416bp) region."
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kundi O.,
RA Laskari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: TO E.COLI PLASMIDS ANTIRESTRICTION PROTEIN KICA/KILC.
CC -1- SIMILARITY: TO E.COLI PLASMIDS ANTIRESTRICTION PROTEIN KICA/KILC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000133; AAC73351.1; -
DR EMBL: D83536; BAA7917.1; ALT_INIT.
DR EMBL: 070214; AAB08668.1; ALT_INIT.
DR Ecogen: EG13336; yafX.
DR InterPro: IPR004914; Antirestrict.
DR Pfam: PF03230; Antirestrict. 1
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 152 AA; 17419 MW; C82749CB8C2C1731 CRC64;
SQ
Query Match 24.1%; Score 53; DB 1; Length 152;
Best Local Similarity 36.4%; Pred. No. 4.7;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;
QY 6 YKRYALKSGSYNAPMPENGQT 27
DB 66 WSEYTLNCGAFMSPEPDNDT 87
RESULT 15
LYTB_STRPN STANDARD; PRT; 658 AA.
ID LYTB_STRPN
AC Q924P7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.36)
DE (Murein hydrolase).
LTYB OR SP0965.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxId=1313;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
RC STRAIN-R6;
RA MEDLINE-99195827; PubMed-10096093;
RA Garcia P., Gonzalez M.P., Garcia F., Lopez R., Garcia J.L.,
RT "LybB, a novel pneumococcal murein hydrolase essential for cell
RT separation."
RL Mol. Microbiol. 31:1275-1281(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4;
RX MEDLINE-21357209; PubMed-11463916;
```

```
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae."
RL Science 293:498-506(2001).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
CC CELL SEPARATION.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC unit in high-mannose glycoproteins and glycoproteins containing
CC the -[Man(GlcNAc)2]asn-structure. One N-acetyl-D-glucosamine
CC residue remains attached to the protein; the rest of the
CC oligosaccharide is released intact.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: A010312; CAA09078.1; -
DR EMBL: AE007400; AAK75086.1; -
DR TIGR: SP0965; -
DR InterPro: IPR002901; Amidase_4.
DR Pfam: PF01832; Amidase_4; 1.
DR SMART: SM00047; LY22; 1.
KW Signal; Hydrolase; Cell wall; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 658
FT FT PUTATIVE ENDO-BETA-N-
FT FT ACETYLGLUCOSAMINIDASE.
FT FT I -> M (IN REF. 1).
FT FT T -> A (IN REF. 1).
FT FT E -> K (IN REF. 1).
FT FT L -> P (IN REF. 1).
FT FT F -> S (IN REF. 1).
FT FT 580
SQ SEQUENCE 658 AA; 76469 MW; B625515006C9CB36 CRC64;
Query Match 24.1%; Score 53; DB 1; Length 658;
Best Local Similarity 38.7%; Pred. No. 24;
Matches 12; Conservative 4; Mismatches 5; Indels 10; Gaps 1;
QY 3 YKNRYALKSGSYNAPMPENGQTEINDWI 33
DB 213 FENGHYLYLKSGGYMAA-----NEWI 233
Search completed: December 10, 2002, 10:54:37
Job time : 7.6 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:49:03 : Search time 21.8 seconds
(without alignments)
378.068 Million cell updates/sec

Title: US-09-142-970-1_COPY_1_40
Perfect score: 220
Sequence: 1 IYKKNRYALKSGSVNAPMENGOTENNDWLMGSTQE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	220	100.0	992	2	057035 neisseria m
2	220	100.0	992	2	0956X5 neisseria m
3	220	100.0	992	2	0956X4 neisseria m
4	220	100.0	992	2	030573 neisseria m
5	220	100.0	997	2	030575 neisseria m
6	220	100.0	1561	2	051169 neisseria m
7	220	100.0	1773	16	093YB9 neisseria m
8	216	98.2	1815	16	093YB4 neisseria m
9	203	92.3	996	2	057309 neisseria m
10	202	91.8	993	2	0956X3 neisseria m
11	202	91.8	996	2	030574 neisseria m
12	202	91.8	997	2	0956X2 neisseria m
13	202	91.8	1552	2	09A1U6 neisseria m
14	101	45.9	1764	2	093Y34 haemophilus
15	59	26.8	221	16	098PJO mycoplasma
16	58.5	26.6	114	16	08XG74 salmonella

17	58.5	26.6	283	9	09AF60	09AF60 streptococc
18	57	25.9	164	2	09REJ1	09REJ1 agrobacteri
19	57	25.9	452	16	08X9H2	08X9H2 salmonella
20	56	25.5	151	12	P89289	P89289 xestia c-ni
21	56	25.5	452	16	08ZRO6	08ZRO6 salmonella
22	56	25.5	452	16	08X921	08X921 escherichia
23	56	25.5	473	16	08RDT2	08RDT2 fusobacteri
24	56	25.5	478	17	097U84	097U84 sulfolobus
25	56	25.5	859	12	09PYP8	09PYP8 xestia c-ni
26	56	25.5	2081	12	09WAL8	09WAL8 sacuna dwa
27	55.5	25.2	175	5	08SW64	08SW64 encephalito
28	55	25.0	162	2	09RGT2	09RGT2 bacteroides
29	55	25.0	214	2	09XD19	09XD19 bacteroides
30	54.5	24.8	600	16	097L33	097L33 clostridium
31	54.5	24.8	948	5	09V900	09V900 drosophila
32	54	24.5	211	10	039489	039489 colocalia e
33	54	24.5	794	10	09LWY7	09LWY7 oryza sativ
34	53.5	24.3	293	16	09RU26	09RU26 deinococcus
35	53.5	24.3	1317	16	0910F4	0910F4 pseudomonas
36	53	24.1	157	2	08VRA1	08VRA1 escherichia
37	53	24.1	159	16	08X3C0	08X3C0 escherichia
38	53	24.1	161	16	08X9K5	08X9K5 escherichia
39	53	24.1	271	2	09AL46	09AL46 shigella fl
40	53	24.1	300	2	059248	059248 bacillus su
41	53	24.1	300	16	P96688	P96688 bacillus su
42	53	24.1	596	5	045633	045633 caenorhabdi
43	53	24.1	614	2	09AHT8	09AHT8 streptococ
44	53	24.1	646	5	021689	021689 caenorhabdi
45	53	24.1	1222	10	094IV6	094IV6 oryza sativ

ALIGNMENTS

RESULT 1
057035 PRELIMINARY; PRT; 992 AA.
ID 057035;
AC 057035;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease precursor (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B40;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Maiorini B., Mueller K., Sella A., Wang J.-F.,
RA del Valle J., Achtman W.,
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
RN [2]
RP SEQUENCE OF 37-532 FROM N.A.
RC STRAIN=HR48;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.,
RA "Comparative characterization of the iga gene encoding Iga1 protease
RA in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae."
RT Mol. Microbiol. 15:495-506(1995).
RL MOL. Microbiol. 15:495-506(1995).
DR EMBL: AF012211; AAC45794.2; -
DR EMBL: X82480; CAA57863.1; -
DR EMBL: X82475; CAA57858.1; -
DR MEMOPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.

```
DR PRINTS; PRO0921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
SQ SEQUENCE 992 AA; 109228 MW; 3677DDEACE69F69 CRC64;

Query Match
Best Local Similarity 100.0%; Score 220; DB 2; Length 992;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
Db 557 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 596

RESULT 2
QY 0956X5 PRELIMINARY; PRT; 992 AA.
AC 0956X5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 19A1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=42491;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012203; AAC45786.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PRO0921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SQ SEQUENCE 992 AA; 109239 MW; 306FF451B7213D49 CRC64;

Query Match
Best Local Similarity 100.0%; Score 220; DB 2; Length 992;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
Db 557 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 596

RESULT 3
QY 0956X4 PRELIMINARY; PRT; 992 AA.
AC 0956X4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 19A1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=423906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012204; AAC45787.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PRO0921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CE3 CRC64;

Query Match
Best Local Similarity 100.0%; Score 220; DB 2; Length 992;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
Db 557 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 596

RESULT 4
QY 030573 PRELIMINARY; PRT; 992 AA.
AC 030573;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 19A1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=423910;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012205; AAC45788.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PRO0921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SQ SEQUENCE 992 AA; 109255 MW; EF7CA563BC7D7ECB CRC64;
```

```
Query Match
Best Local Similarity 100.0%; Score 220; DB 2; Length 992;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 40
Db 557 LYKKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 596
```

```
RESULT 5
030575
ID 030575 PRELIMINARY; PRT; 997 AA.
AC 030575;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24099;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012210; AAC45793.2; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin-sup.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
FT NON_TER 1
FT NON_TER 997
SQ SEQUENCE 997 AA; 110152 MW; 4DE40DE594E5E28E CRC64;

Query Match 100.0%; Score 220; DB 2; Length 997;
Best Local Similarity 100.0%; Pred. No. 5.7e-21;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 40
Db 558 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 597

RESULT 6
051169
ID 051169 PRELIMINARY; PRT; 1561 AA.
AC 051169;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease.
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF13;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the Iga gene encoding Iga1 protease
  in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
  influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82474; CAA57857.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin-sup.
DR Pfam; PF02395; IGA1; 1.
```

```
DR Pfam; PF03212; Pertactin; 2.
DR PRINTS; PR00921; IGASERPASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1561 AA; 171849 MW; 1C96E291A00017B5 CRC64;

Query Match 100.0%; Score 220; DB 2; Length 1561;
Best Local Similarity 100.0%; Pred. No. 9.7e-21;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 40
Db 584 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 623

RESULT 7
090VB9
ID 090VB9 PRELIMINARY; PRT; 1773 AA.
AC 090VB9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease (EC 3.4.21.7).
GN IGA OR NMA0905.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
  Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
  Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
  Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
  Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
  Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
  meningitidis 22491.";
RL Nature 404:502-506(2000).
DR EMBL; AL162754; CAB84182.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin-sup.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease; Hydrolase; Complete proteome.
SQ SEQUENCE 1773 AA; 196350 MW; CAC19E7313D76CE1 CRC64;

Query Match 100.0%; Score 220; DB 16; Length 1773;
Best Local Similarity 100.0%; Pred. No. 1.1e-20;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 40
Db 573 LYKKNRYVALKSGGSYNAPMPENGQTEENDWILMGSTOE 612

RESULT 8
09KOB4
ID 09KOB4 PRELIMINARY; PRT; 1815 AA.
AC 09KOB4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga-specific serine endopeptidase.
GN NMB0700.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
```

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=2017575; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scariato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
DR EMBL: AE002424; AAF4117.1; -.
DR MEROPS: S06.001; -.
DR TIGR: NMB0700; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 1815 AA; 201077 MW; 2259D4D7176ZC57F CRC64;
```

```

Query Match          98.2%; Score 216; DB 16; Length 1815;
Best Local Similarity 97.5%; Pred. No. 4.1e-20;
Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 1 LYYKNRYVALKSGGSYNAPMENGOTENNNDWILMGSTOE 40
DB 573 LYYKNRYVALKSGGSYNAPMENGOTENNNDWILMGSTOE 612

RESULT 9
ID 057309 PRELIMINARY; PRT; 996 AA.
AC 057309;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE IGA1 protease precursor (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z3524;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achtmann M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
RN [2]
RP SEQUENCE OF 37-532 FROM N.A.
RC STRAIN=ETH2;
RX MEDLINE=95302961; PubMed=783620;
RA Lombolt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IGA1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae."
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL: AF012207; AAC45790.2; -.
DR EMBL: X82469; CAA57851.1; -.
DR EMBL: X82468; CAA57851.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
```

```

DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 996
SQ SEQUENCE 996 AA; 109717 MW; AB39F7B98A41F986 CRC64;
```

```

Query Match          92.3%; Score 203; DB 2; Length 996;
Best Local Similarity 90.0%; Pred. No. 1.1e-18;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 LYYKNRYVALKSGGSYNAPMENGOTENNNDWILMGSTOE 40
DB 557 LYYKNRYVALKSGGSYNAPMENGOTENNNDWVFMGTYOE 596
```

```

RESULT 10
ID 09S6X3 PRELIMINARY; PRT; 993 AA.
AC 09S6X3;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z4400;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achtmann M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012208; AAC45791.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 993
SQ SEQUENCE 993 AA; 109441 MW; 109FAA2EF88AC3C6 CRC64;

Query Match          91.8%; Score 202; DB 2; Length 993;
Best Local Similarity 90.0%; Pred. No. 1.5e-18;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```

QY 1 LYYKNRYVALKSGGSYNAPMENGOTENNNDWILMGSTOE 40
DB 558 LYYKNRYVALKSGGSYNAPMENGVTENNNDWVFMGTYOE 597

RESULT 11
ID 030574 PRELIMINARY; PRT; 996 AA.
AC 030574;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
```

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24024;
RX MEDLINE=98010345; PubMed=9350862;
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR MEROPS; S06.001; -
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER
KW NON_TER
SQ SEQUENCE 996 AA; 109688 MW; B3ABDEF5E54C337D9 CRC64;

Query Match
Best Local Similarity 91.8%; Score 202; DB 2; Length 996;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPENGOTENNNDWILMGSTOE 40
Db 557 LYKKNRYVALKSGSVNAPENGVTENNNDWVFMGYTOE 596

RESULT 12
Q9S6X2 PRELIMINARY; PRT; 997 AA.
AC Q9S6X2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24081;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012209; AAC45792.2; -
DR MEROPS; S06.001; -
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER
KW NON_TER
SQ SEQUENCE 997 AA; 109812 MW; 06F2E361E7E202ED CRC64;

Query Match
Best Local Similarity 91.8%; Score 202; DB 2; Length 997;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPENGOTENNNDWILMGSTOE 40
Db 558 LYKKNRYVALKSGSVNAPENGVTENNNDWVFMGYTOE 597

RESULT 13
Q9A1U6 PRELIMINARY; PRT; 1552 AA.
AC Q9A1U6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Iga1 protease.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RA Vltovskii S., Savers J.R.;
RT "degenerate specificity of Neisseria meningitidis Iga1 protease."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF235032; AAK15023.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 2.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
SQ SEQUENCE 1552 AA; 171205 MW; E49D6657B120CEFF CRC64;

Query Match
Best Local Similarity 91.8%; Score 202; DB 2; Length 1552;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGSVNAPENGOTENNNDWILMGSTOE 40
Db 574 LYKKNRYVALKSGSVNAPENGVTENNNDWVFMGYTOE 613

RESULT 14
Q93T34 PRELIMINARY; PRT; 1764 AA.
AC Q93T34;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Iga1 protease type 2.
GN IGA1.
OS Haemophilus aegyptius.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F3031;
RA McGillivray G., Actis L.A.;
RT "Iga protease from H. aegyptius F3031."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF369907; AAK56925.1; -
DR MEROPS; S06.001; -
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGA1; 1.
DR Pfam; PF03212; Pertactin; 1.
KW Protease.
SQ SEQUENCE 1764 AA; 193863 MW; EC583CDEB81DBE6C CRC64;

Query Match
Best Local Similarity 45.9%; Score 101; DB 2; Length 1764;
Matches 40.5%; Pred. No. 0.00014;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:14:59 ; Search time 28.2 Seconds
(without alignments)
189.008 Million cell updates/sec

Title: US-09-142-970-2_COPY_1_40
Perfect score: 224
Sequence: 1 LYXKNRYALKSGSVNAPMPENGVTENNDFVEMGYTQE 40

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	224	100.0	104	19	AAW61603
2	212	94.6	104	19	AAW61604
3	211	94.2	104	19	AAW61605
4	210	93.8	104	19	AAW61606
5	202	90.2	105	19	AAW65656
6	202	90.2	105	19	AAW61602
7	98	43.8	1541	11	AAW61602
8	58	25.9	452	22	AAU34434
9	58	25.9	452	22	AAU38481
10	58	25.9	452	22	AAW61602

11	56	25.0	359	22	AAW51254
12	55	24.6	108	22	AAO10304
13	54.5	24.3	971	21	AAV95687
14	54.5	24.3	1364	22	AAW70912
15	53.5	23.9	194	23	ABP07412
16	53.5	23.9	456	17	AAW98365
17	53.5	23.9	459	22	ABW71664
18	53.5	23.9	569	19	AAW61214
19	53.5	23.9	569	23	ABP54633
20	53.5	23.9	591	19	AAW55099
21	53.5	23.9	591	23	ABP54593
22	53.5	23.9	648	16	AAW78933
23	53.5	23.9	648	18	AAW08967
24	53.5	23.9	648	19	AAW54125
25	53.5	23.9	648	21	AAW51781
26	53.5	23.9	648	21	AAW80363
27	53.5	23.9	659	21	AAW81517
28	53.5	23.9	671	17	AAW85290
29	53.5	23.9	678	21	AAW81667
30	52.5	23.4	1577	17	AAW91047
31	52	23.2	417	15	AAW47873
32	52	23.2	417	18	AAW03594
33	52	23.2	417	20	AAW97238
34	52	23.2	417	20	AAW89243
35	52	23.2	417	22	AAU02431
36	52	23.2	695	22	AAW93314
37	52	23.2	695	23	ABW97385
38	52	23.2	1061	22	ABW65553
39	51.5	23.0	213	23	ABW54972
40	51.5	23.0	512	6	AAW50629
41	51.5	23.0	578	20	AAW09065
42	51.5	23.0	948	22	ABW58055
43	51	22.8	109	21	ABW41000
44	51	22.8	109	23	ABP32003
45	51	22.8	208	22	AAW63264

ALIGNMENTS

RESULT 1
AAW61603
AAW61603 standard; peptide; 104 AA.
AC AAW61603;
DT 27-OCT-1998 (first entry)
DE Neisseria Igal protease fragment 2.
KW Immunoglobulin protease; carrier; paediatric; vaccine;
epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
OS Neisseria sp.
XX W09831791-A1.
XX 23-JUL-1998.
XX 20-JAN-1998; 98WO-EP00294.
XX 21-JAN-1997; 97EP-0100883.
XX (PLAC) MAX PLANK GES FOERDERUNG WISSENSCHAFTEN.
XX (INNR) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX Achtmann M, Moreau M;
XX WPT: 1998-414092/35.
XX New peptide from Neisseria immunoglobulin protease - useful as
XX immunogenic carrier, e.g. particularly for polysaccharide(s),
XX forming conjugates used in vaccines against Neisseria and
XX

```

PT Haemophilus
XX
XX Claim 2; Fig 2; 32pp; English.
PS
CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 104 AA;
SQ
Query Match 100.0%; Score 224; DB 19; Length 104;
Best Local Similarity 100.0%; Pred. No. 3.6e-23;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40
Db 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40

RESULT 2
AAW61604 standard; peptide; 104 AA.
ID AAW61604
XX
XX AAW61604;
AC
XX
XX 27-OCT-1998 (first entry)
DT
XX
XX Neisseria Iga1 protease fragment 3.
DE
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX
XX WO9831791-A1.
PN
XX
XX 23-JUL-1998.
PD
XX
XX 20-JAN-1998; 98WO-EP00294.
PF
XX
XX 21-JAN-1997; 97EP-0100883.
PR
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
PI
XX
XX WPI: 1998-414092/35.
DR
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 2; Fig 1; 32pp; English.
PS
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 104 AA;
SQ
Query Match 94.6%; Score 212; DB 19; Length 104;
Best Local Similarity 92.5%; Pred. No. 2.2e-21;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

OY 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40
Db 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40

RESULT 3
AAW61605 standard; peptide; 104 AA.
ID AAW61605
XX
XX AAW61605;
AC
XX
XX 27-OCT-1998 (first entry)
DT
XX
XX Neisseria Iga1 protease fragment 4.
DE
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX
XX WO9831791-A1.
PN
XX
XX 23-JUL-1998.
PD
XX
XX 20-JAN-1998; 98WO-EP00294.
PF
XX
XX 21-JAN-1997; 97EP-0100883.
PR
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
PI
XX
XX WPI: 1998-414092/35.
DR
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 2; Fig 1/4; 32pp; English.
PS
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 104 AA;
SQ
Query Match 94.2%; Score 211; DB 19; Length 104;
Best Local Similarity 92.5%; Pred. No. 2.2e-21;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40
Db 1 LYYKNRYALKSGSVNAPMPENGVTENNNDVFMGYTOE 40

RESULT 4
AAW61606 standard; peptide; 104 AA.
ID AAW61606
XX
XX AAW61606;
AC
XX
XX 27-OCT-1998 (first entry)
DT

```


DE Neisseria IgA1 protease fragment 5.
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX WO9831791-A1.
PN
XX 23-JUL-1998.
PD
XX 20-JAN-1998; 98WO-EP00294.
PF
XX 21-JAN-1997; 97EP-0100883.
PR
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
PI
XX WPI: 1998-414092/35.
DR
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 2; Fig 4; 32pp; English.
PS
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
CC
XX Sequence 104 AA:

Query Match 93.8%; Score 210; DB 19; Length 104;
Best Local Similarity 90.0%; Pred. No. 3.1e-21;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LYYKNRYRYALKSGSVNAPMENGVTENNNDWVFMGYTOE 40
1 LYYKNRYRYALKSGGRUNAPMENGVAENNDWFMGYTOE 40
DB

RESULT 5
AAW65656
ID AAW65656 standard; peptide; 105 AA.
XX
XX AAW65656;
AC
XX
XX 15-OCT-1998 (first entry)
DT
XX
XX 105-mer peptide used in polysaccharide-peptide conjugate.
DE
XX
XX Polysaccharide-peptide conjugate; vaccine; linker moiety; adjuvant;
KW immune response.
XX
XX Synthetic.
OS
XX
XX Key Location/Qualifiers
FH Modified-site 1 /note= "N-terminal acetyl"
FT
XX
XX WO9831393-A2.
PN
XX 23-JUL-1998.
PD
XX 21-JAN-1998; 98WO-EP00654.
PR

XX
XX 21-JAN-1997; 97EP-0100884.
PR
XX
XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA
XX
XX Mistretta N, Moreau M;
PI
XX
XX WPI: 1998-413820/35.
DR
XX
XX Polysaccharide-peptide conjugates, useful as, e.g. vaccines -
PT comprise peptide moiety with at least six amino acid residues,
PT polysaccharide chain with at least four repeat units, and linker
PT moiety
XX
XX Example 1; Page 14; 28pp; English.
PS
XX The invention relates to: (A) polysaccharide-peptide conjugate (in which
CC the polysaccharide is immunogenic), comprising: (a) a peptide moiety
CC which has at least 6 amino acid residues, at least 1 of which is a
CC cysteine residue; (b) a polysaccharide chain comprising at least 4 repeat
CC units, and (c) a linker moiety bound to the thiol group of the cysteine.
CC The linker is also bound to: (i) native amino, hydroxyl or carboxyl
CC groups of the polysaccharide chain; (ii) amino groups created by
CC hydrolysis of native N-acetyl groups of the polysaccharide chain, or (iii)
CC functional groups introduced on the polysaccharide chain upon
CC derivatisation with a spacer moiety bound to native amino, hydroxyl or
CC carboxyl groups of the polysaccharide chain, and (v) conjugating a
CC peptide as in (Aa) to a polysaccharide chain comprising at least 4 repeat
CC units, comprising: (a) coupling the peptide to a linker through the thiol
CC group of the cysteine residue, and (b) coupling the linker to the
CC polysaccharide chain through one of groups (i), (ii) or (iii) as
CC described in (A). The conjugates are especially useful as vaccines to
CC elicit a protective long term immune response against a pathogenic
CC microorganism from which the immunogenic polysaccharide chain is derived.
CC Known polysaccharide-peptide conjugates are less immunogenic than their
CC polysaccharide-protein counterparts and require adjuvantation. The new
CC conjugates have good immunogenicity, largely because the peptides are of
CC sufficient size. The present sequence represents a peptide used in
CC a polysaccharide-peptide conjugate.
XX
XX Sequence 105 AA:

Query Match 90.2%; Score 202; DB 19; Length 105;
Best Local Similarity 90.0%; Pred. No. 4e-20;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 LYYKNRYRYALKSGSVNAPMENGVTENNNDWVFMGYTOE 40
2 LYYKNRYRYALKSGSVNAPMENGVTENNNDWFMGYTOE 41
DB

RESULT 6
AAW61602
ID AAW61602 standard; peptide; 105 AA.
XX
XX AAW61602;
AC
XX
XX 27-OCT-1998 (first entry)
DT
XX
XX Neisseria IgA1 protease fragment 1.
DE
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX
XX WO9831791-A1.
PN
XX 23-JUL-1998.
PD
XX 20-JAN-1998; 98WO-EP00294.
PF
XX 21-JAN-1997; 97EP-0100883.
PR


```
CC ftp.wipo.int/pub/published_pct_sequences.  
XX  
SO Sequence 452 AA;  
  
Query Match 25.9%; Score 58; DB 22; Length 452;  
Best Local Similarity 38.5%; Pred. NO. 18;  
Matches 10; Conservative 7; Mismatches 5; Indels 4; Gaps 1;  
  
Oy 12 KSGSVNAPMPENGV----TENNDVV 33  
      :|:::|||||:||||:  
Db 200 KAKGEIFGSLPENGIAIMNADNNMWL 225  
  
RESULT 9  
AAU38481  
ID AAU38481 standard; Protein: 452 AA.  
XX  
AC AAU38481;  
XX  
DT 14-FEB-2002 (first entry)  
DE Salmomella typhi cellular proliferation protein #372.  
XX  
KW Antisense; prokaryotic cellular proliferation protein;  
KM antibiotic; antibacterial; drug design.  
OS Salmomella typhi.  
XX  
PN WO200170955-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 21-MAR-2001; 2001WO-US09180.  
PE  
PR 21-MAR-2000; 2000US-191078P.  
PR 23-MAY-2000; 2000US-206848P.  
PR 26-MAY-2000; 2000US-207727P.  
PR 23-OCT-2000; 2000US-242578P.  
PR 27-NOV-2000; 2000US-253625P.  
PR 22-DEC-2000; 2000US-257931P.  
PR 16-FEB-2001; 2001US-269308P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,  
PI Yamamoto RT, Xu HH;  
DR WPI: 2001-611495/70.  
DR N-PSTB: AAS56340.  
PS  
PT New polynucleotides for the identification and development of  
PT antibiotics, comprise sequences of antisense nucleic acids -  
XX  
XX Example 3; Seq ID No 14074; 511pp; English.
```

CC	ftp.wipo.int/pub/published_pct_sequences.
XX	
SQ	Sequence 452 AA;
	Query Match 25.9%; Score 58; DB 22; Length 452; Best Local Similarity 38.5%; Pred. No. 18; Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1
OY	12 KSGGSVNAAPMPCNV---TENNDWV 33 1 : I : : : : : II : : : : : DB 200 KAKGEIYTGLPENGIAIMNADNNMDL 225
RESULT 10	
ID	AAG98402 standard; Protein; 452 AA.
XX	AAG98402
AC	AAG98402;
XX	
DT	21-SEP-2001 (first entry)
XX	
DE	Escherichia coli protein sequence SEQ ID NO:450.
XX	
KW	Escherichia coli; identification; proliferation; microorganism;
KW	antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;
KW	bacterial growth inhibition.
XX	
OS	Escherichia coli.
XX	
PN	WO200148209-A2.
PD	05-JUL-2001.
PF	19-DEC-2000; 2000WO-US34419.
PR	23-DEC-1999; 99US-0173005.
PA	(ELIT-) ELITRA PHARM INC.
PI	Forsyth RA, Ohlsen KL, Zyskind JW;
DR	WPI: 2001-457376/49.
N-P	PSDB: AAH81458.
PT	Novel nucleic acids encoding proteins required for Escherichia coli
XX	proliferation, useful for screening for antimicrobial agents -
PS	Claim 19; Page 569; 596pp: English.
XX	
CC	The present invention describes a purified or isolated nucleic acid
CC	sequence (I) consisting essentially of one of the 93 nucleotide sequences
CC	given in AAH81202 to AAH81294, where expression of the nucleic acid in a
CC	microorganism is capable of inhibiting proliferation of a microorganism.
CC	(I) have antibacterial and antibiotic activities, and can be used in
CC	gene therapy. Expression of (I) in a microorganism inhibits proliferation
CC	of the microorganism, and the manufactured antibiotic is useful for
CC	reducing the activity or level of a gene product required for
CC	proliferation of a microorganism in a subject, specifically humans. The
CC	nucleic acids that inhibit bacterial growth or proliferation can be used
CC	as antisense therapeutics for killing bacteria. In addition to
CC	therapeutic applications, the nucleic acid sequences complementary to
CC	sequences required for proliferation can be used as diagnostic tools.
CC	For example, nucleic acid probes complementary to proliferation-required
CC	sequences that are specific for particular species of microorganisms can
CC	be used as probes to identify particular microorganism species in
CC	clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli
CC	proteins given in AAG98239 to AAG98431, and AAH81488 to AAH81491
CC	represent oligonucleotides, which are used in the exemplification of the
XX	present invention.
SQ	Sequence 452 AA;

Best Local Similarity 38.5%; Pred. No. 18;
Matches 10; Conservative 7; Mismatches 5; Indels 4; Gaps 1;

OY 12 KSGGVNAPMPENGVT---TENNDMW 33
Db 200 KAKEIFSGLEPENGIAIMADNDWL 225

RESULT 11

AAB51254
ID AAB51254 standard; Protein; 359 AA.

AC AAB51254;

DT 27-MAR-2001 (first entry)

DE Human CD72 protein sequence SEQ ID NO:7.

KW CD100; CD72; screening; viral infection; bacterial infection; cancer;
KM fungal infection; infectious disease; allergy; autoimmune disease.

OS Homo sapiens.

PN WO200075655-A1.

PD 14-DEC-2000.

PF 01-JUN-2000; 2000WO-JP03558.

PR 03-JUN-1999; 99JP-0157111.

PA (TAKE) TAKEDA CHEM IND LTD.

PI Kikuchi H, Kumanogoh A, Hori A;

DR WPI: 2001-061765/07.

DR N-PSDB; AAB51254.

PT Screening of compounds modifying the binding of CD100 to CD72 for use
in treatment of infectious diseases, cancer and disorders of antibody
production.

PS Disclosure; Page 105-106; 110pp; Japanese.

CC The present invention describes a method of screening for compounds
which modify the binding of CD100 and its salts to CD72 and its salts.

CC CD100 is contacted with CD72 in the presence and absence of the compound
and the degree of binding is compared. Also described are: (1) kits for

CC carrying out the novel method; (2) compounds identified by the novel
method; (3) drug compositions containing the compounds of (2);

CC (4) non-human CD100 knockout animals; (5) screening compounds as drugs
for the prevention and treatment of CD100-associated diseases, using the

CC knockout animals of (4); (6) transgenic non-human animals transformed
with a CD100 gene or modified CD100 gene; and (7) screening methods

CC using the transgenic animals of (6). The method can be used for
identifying compounds for the treatment and prevention of bacterial,

CC viral and fungal infections, cancers, and diseases involving abnormal
antibody production or excess antibody production, such as allergies and

CC autoimmune diseases. The present sequence represents human CD72 which is
given in the exemplification of the present invention.

XX Sequence 359 AA;

Query Match 25.0%; Score 56; DB 22; Length 359;

Best Local Similarity 34.9%; Pred. No. 25;
Matches 15; Conservative 7; Mismatches 15; Indels 6; Gaps 2;

OY 1 LYKKNRYVALKS---GGSVNAAPMPENGVTENNDWVFNGYTO 39

Db 274 IYQSHSYFLNLSLPLNGSGNSY--TGLSSNKDKMLDDDTQ 314

RESULT 12

AA010304
ID AA010304 standard; Protein; 108 AA.

AC AA010304;

DT 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 24196.

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;
nervous system disorders; arthritis; inflammation.

OS Homo sapiens.

PN WO200164835-A2.

PD 07-SEP-2001.

PF 26-FEB-2001; 2001WO-US04927.

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Drmanac RT;

DR WPI: 2001-514838/56.

DR N-PSDB; AA190235.

PT Isolated nucleic acids and polypeptides, useful for preventing
diagnosing and treating e.g. leukaemia, inflammation and immune
disorders -

PS Claim 20; SEQ ID NO 24196; 1399pp + Sequence Listing; English.

CC The invention relates to human polynucleotides (AA179941-AA193841) and
the encoded proteins (AA000010-AA013910) that exhibit activity elating to

CC cytokine, cell proliferation or cell differentiation or which may induce
production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating
activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or
treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 108 AA;

Query Match 24.6%; Score 55; DB 22; Length 108;

Best Local Similarity 44.0%; Pred. No. 8;
Matches 11; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

OY 15 GSVNAPMPENGVTENNDWVFNGYTO 39

Db 4 GVPGAPLSPGVTPEHKOMFFRPYTO 28

RESULT 13

AA195687
ID AA195687 standard; Protein; 971 AA.

AC AA195687;

DT 25-OCT-2000 (first entry)

DE Cosmid cHRIM5 encoded protein p2-0f.

```

XX  Cosmid cHRIM5; nematocidae; nematode; biological control agent;
KW  transgenic plant; helminthiasis; P2-Of.
XX
XX  Xenorhabdus bovienii.
XX
XX  WO200042855-A1.
XX
XX  27-JUL-2000.
XX
XX  24-JAN-2000; 2000WO-GB00219.
XX
XX  22-JAN-1999; 99GB-0001499.
XX
XX  (HORT-) HORTICULTURE RES INT.
XX
XX  Morgan JAW, Jarrett P, Ellis D, Ousley MA:
XX
XX  WPI: 2000-499157/44.
XX  N-PSDB: AAA50029.
XX
XX  Novel composition used to control parasitic nematodes, especially in
XX  plants such as maize, cotton, soya, and rice, comprises a bacterium
XX  which is a symbiont of an entomopathogenic nematode -
XX
XX  Example 6; Page 38-39; 74pp: English.
XX
XX  The present sequence is that of protein P2-Of encoded by an open
XX  reading frame identified in cosmid cHRIM5 (see AAA50029). cHRIM5 was
XX  obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986)
XX  SauA-digested DNA fragments into the BamHI site of the Stratagene
XX  cosmid vector Supercos1, packaging into Escherichia coli XL Blue 1,
XX  and screening for nematocidal activity against Caenorhabditis elegans.
XX  Analysis of the DNA indicated a number of open reading frames for
XX  which the corresponding protein sequences were determined (see
XX  AA155665-155735). Nematodes can be controlled through the use of
XX  bacteria associated symbiotically with an entomopathogenic nematode.
XX  Such bacteria include Xenorhabdus and Photorhabdus spp. such as X.
XX  bovienii strain I73. The symbiont bacteria, an engineered
XX  bacterium, or a nematocidal protein obtained from such bacteria,
XX  can be used to control helminthiasis in a human or domesticated
XX  animal or for the control of plant pathogen nematodes. Also
XX  claimed are vectors for expressing nematocidal proteins in host
XX  cells, and transgenic plants.
XX
XX  Sequence 971 AA:
XX
XX  Query Match 24.3%; Score 54.5; DB 21; Length 971;
XX  Best Local Similarity 27.9%; Pred. No. 1.4e+02;
XX  Matches 17; Conservative 7; Mismatches 16; Indels 21; Gaps 3;
XX
XX  QY 1 LYYKXRYVALKSG-----GSV-----NAPM---PENGVTEENNDFVFGTYQ 39
XX  DB 654 LYYGRTYQPMAGSWLSADPAGTIDGLNLRYMRNRPATLDDKGLADCGNRYFFPFLH 713
XX
XX  QY 40 E 40
XX  DB 714 E 714
XX
XX  RESULT 14
XX  AAG70912
XX  AAG70912 standard; Protein; 1364 AA.
XX
XX  AAG70912;
XX
XX  27-JUL-2001 (first entry)
XX
XX  S cerevisiae apoptosis associated protein YGR183C.
XX
XX  Yeast; fungus; apoptosis; infection; proliferative disease;
XX  vaccine; autoimmune disease; ischaemia; neurodegeneration.
XX

```

```

OS  Saccharomyces cerevisiae.
XX
XX  WO200102550-A2.
XX
XX  11-JAN-2001.
XX
XX  03-JUL-2000; 2000WO-BE00077.
XX
XX  01-JUL-1999; 99EP-0870141.
XX
XX  (JANC ) JANSSEN PHARM NV.
XX
XX  Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
XX  Pi Nelissen BJM, Reekmans RJ;
XX
XX  WPI: 2001-367042/38.
XX  N-PSDB: AAH29948.
XX
XX  Yeast and fungal nucleic acids encoding proteins involved in a pathway
XX  leading to programmed cell death, useful for treating proliferative
XX  disorders, yeast and fungal infections, or for preventing apoptosis in
XX  certain diseases -
XX
XX  Claim 1; Fig 1; 218pp: English.
XX
XX  The present invention provides the protein and coding sequences of a
XX  number of apoptosis associated proteins from the yeast Saccharomyces
XX  cerevisiae and the fungus Candida albicans. These can be used to identify
XX  treatments for fungal and yeast infections, for proliferative diseases
XX  and for apoptosis related diseases such as autoimmune diseases, ischaemia
XX  and neurodegeneration. The present sequence is one of the S. cerevisiae
XX  proteins of the invention.
XX
XX  Sequence 1364 AA:
XX
XX  Query Match 24.3%; Score 54.5; DB 22; Length 1364;
XX  Best Local Similarity 37.9%; Pred. No. 2.1e+02;
XX  Matches 11; Conservative 6; Mismatches 7; Indels 5; Gaps 1;
XX
XX  QY 3 YKNRYRYVALKSGSVNAPMPENCVTENNND 31
XX  DB 519 YRSHRFSIRKSGSVG----NSNTINGND 542
XX
XX  RESULT 15
XX  ABP07412
XX  ID ABP07412 standard; Protein; 194 AA.
XX
XX  AC ABP07412;
XX
XX  24-JUN-2002 (first entry)
XX
XX  Human OREFX protein sequence SEQ ID NO:14806.
XX
XX  DE Human; Open reading frame; OREFX; gene therapy; cancer; cirrhosis;
XX  KW hyperproliferative disorder; psoriasis; benign tumour; hemorrhage;
XX  KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
XX  KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
XX  KW hypertension; hypothyroidism; cholesterol ester storage disease;
XX  KW immune deficiency; immune disorder; infectious disease;
XX  KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
XX  KW myasthenia gravis.
XX
XX  OS Homo sapiens.
XX
XX  PN WO200192523-A2.
XX
XX  PD 06-DEC-2001.
XX
XX  PF 29-MAY-2001; 2001WO-US10836.
XX
XX  PR 30-MAY-2000; 2000US-206132P.
XX  PR 29-AUG-2000; 2000US-228716P.
XX

```

XX (CURA-) CURAGEN CORP.
 PA
 XX
 PI Shinkets RA, Leach MD;
 XX
 DR WPI, 2002-106308/14.
 DR N-PSDB; ABN23164.
 XX

PT Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders
 XX

PS Disclosure; SEQ ID 14806; 1037pp; English.
 XX

CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX
 CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
 CC treating or preventing a pathology associated with an ORFX-associated
 CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumors, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage.
 CC N.B. The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SO Sequence 194 AA;

QY Query Match 23.9%; Score 53.5; DB 23; Length 194;
 Best Local Similarity 44.8%; Pred. No. 26;
 Matches 13; Conservative 3; Mismatches 12; Indels 1; Gaps 1;
 OY 4 KNRYYALKS-GGSVNAPEMPCVTEPENNND 31
 DB 85 EDLYTNKRKSGENKPESEKETAEENNND 113

Search completed: December 10, 2002, 10:54:03
 Job time : 29.2 secs


```

? PRIOR APPLICATION NUMBER: US 60/216593
? PRIOR FILING DATE: 2000-07-07
? NUMBER OF SEQ ID NOS: 17
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO. 15
? LENGTH: 421
? TYPE: PRT
? ORGANISM: Saccharomyces
? US-09-901-252-15

```

Query Match	27.9%	Score 62.5;	DB 10;	Length 421;
Best Local Similarity	32.4%;	Pred. No. 0.46;		
Matches 12; Conservative	11;	Mismatches 13;	Indels 1;	Gaps 1

Oy 3 YKNRYTALKSGG-SVNAPMPENGTYENNDVPMGYT 38
 |::|::|||::|::|::|::
Db 384 YKHFTYLRVFNGCHMVPFDVPENALSMVNEWTIHGGFS 420

```

RESULT 3
US-09-741-669-450
: Sequence 450, Application US/09741669
: Patent No. US20020022718A1
: GENERAL INFORMATION:
: APPLICANT: Forsyth, R. Allyn
: APPLICANT: Ohlssen, Kari L.
: APPLICANT: Zyskind, Judith W.
: TITLE OF INVENTION: Genes identified as required for
: proliferation of E. coli
: FILE REFERENCE: ELITRA.009A
: CURRENT FILING DATE: 2000-12-19
: PRIOR FILING DATE: 1999-12-23
: PRIOR APPLICATION NUMBER: US 60/173005
: NUMBER OF SEQ ID NOS: 481
: SOFTWARE: FASTSEQ for Windows Version 4.0
: SEQ ID NO 450
: LENGTH: 452
: TYPE: PRT
: ORGANISM: Escherichia coli .
US-09-741-669-450

```

Query Match	25.9%	Score 58;	DB 10;	Length 452;	.
Best Local Similarity	38.5%;	Pred. No. 2;			
Matches 10;	Conservative 7;	Mismatches 5;	Indels 4;	Gaps 1	

```
.QY      12 KSGGSVNAMPENGV-----TENNDWV 33  
          | : | : : ||| : ||| :  
Db       200 KAKGEIFSGLPENGIAMNADNDWL 225
```

RESULT 4
US-09-815-242-10027
Sequence 10027, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlssen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23

```

? PRIOR APPLICATION NUMBER: 60/207,727
? PRIOR FILING DATE: 2000-05-26
? PRIOR APPLICATION NUMBER: 60/242,578
? PRIOR FILING DATE: 2000-10-23
? PRIOR APPLICATION NUMBER: 60/253,625
? PRIOR FILING DATE: 2000-11-27
? PRIOR APPLICATION NUMBER: 60/257,931
? PRIOR FILING DATE: 2000-12-22
? PRIOR APPLICATION NUMBER: 60/269,308
? PRIOR FILING DATE: 2001-02-16
? NUMBER OF SEQ ID NOS: 14110
? SOFTWARE: FASTSEQ for Windows Version 4.0
? SEQ ID NO 10027
? LENGTH: 452
? TYPE: prt
? ORGANISM: Escherichia coli
US-09-815-242-10027

```

Query Match	25.9%	Score 58;	DB 10;	Length 452;
Best Local Similarity	38.5%;	Pred. No. 2;		
Matches 10; Conservative	7;	Mismatches 5;	Indels 4;	Gaps 1;

```
QY      12 KSGGSVNAPMPENG-----TENNDWV  33
        | : | : : | | | : | | | :
Db      200 KAKGEIFSGLPENGIAIMNADNDWL  225
```

RESULT 5
US-09-815-242-14074
; Sequence 14074, Application US/09815242
; Patent No. US20020061569A1

: APPLICANT: Haselbeck, Robert
 : APPLICANT: Ohlsen, Kari L.
 : APPLICANT: Zyskind, Judith W
 : APPLICANT: Wall, Daniel
 : APPLICANT: Trawick, John D.
 : APPLICANT: Carr, Grant J.
 : APPLICANT: Yamamoto, Robert T

APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in

```

; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242

```

;; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078

;; PRIOR FILING DATE: 2000-03-21
;; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
 ; PRIOR APPLICATION NUMBER: 60/207,727

; PRIOR FILING DATE: 2000-05-26
 ; PRIOR APPLICATION NUMBER: 60/242,578

;; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-10-23

; PRIOR FILING DATE: 2000-11-27
 ; PRIOR APPLICATION NUMBER: 60/257,931

PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308

PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

```

SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1A07A

```

TYPE: DBT
; LENGTH: 452
; SEX ID NO 140

ORGANISM: *Salmonella typhi*

```

;
FEATURE;
NAME/KEY: VARIANT
;

```

```

; LOCATION: (1)..(452)
; OTHER INFORMATION: Xaa = Any Amino Acid

```

Query Match	25.98;	Score 58;	DB 10;	Length 452;
Best Local Similarity	38.58;	Pred. NO. 2;		

Query Match	25.9%;	Score 58;	L
Best Local Similarity	38.5%;	Pred. No. 2;	

Query Match	25.9%;	Score 58;	L
Best Local Similarity	38.5%;	Pred. No. 2;	

Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;
QY 12 KSGSVNAPMPGV-----TENNDDV 33
1: | : |||| : |||| :
Db 200 KAKGEIYTGIPENGIAIMNNDNL 225

RESULT 6
US-09-765-272-154
Sequence 154, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 154:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 154:

US-09-765-272-154
Query Match 23.9%; Score 53.5; DB 10; Length 569;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

QY 3 YKNRYVALKSGSVNAP-----MPENGVTENNDDVFMGYTQ 39
1: | : |||| : |||| :
Db 129 FENGHYIYKSGGYMANEMIMDKESWFLKPGKMAEKEMVYDHSQ 176

RESULT 7
US-09-765-272-74
Sequence 74, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESS: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland

US-09-765-272-74
Query Match 22.8%; Score 51; DB 10; Length 3075;
Best Local Similarity 28.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 5; Mismatches 17; Indels 14; Gaps 2;

QY 5 NYRYVALKSGSV-----NAPMPENGVTENNDDVFMGYTQ 40
1: | : |||| : |||| :
Db 1253 NEPPQVLINGIRIKROYITMDAPAPENGVRQDEVMARENFKYFNSVSE 1302

RESULT 9

COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:

US-09-765-272-74
Query Match 23.9%; Score 53.5; DB 10; Length 591;
Best Local Similarity 27.1%; Pred. No. 11;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

QY 3 YKNRYVALKSGSVNAP-----MPENGVTENNDDVFMGYTQ 39
1: | : |||| : |||| :
Db 190 FENGHYIYKSGGYMANEMIMDKESWFLKPGKMAEKEMVYDHSQ 237

RESULT 8
US-09-938-275-5
Sequence 5, Application US/09938275
Patent No. US20020111309A1
GENERAL INFORMATION:
APPLICANT: Gerardo Casti1110
APPLICANT: Alan Snow
TITLE OF INVENTION: Therapeutic and Diagnostic Applications
of Laminin and Laminin-Derived Protein Fragments
FILE REFERENCE: PROTEO.P03
CURRENT APPLICATION NUMBER: US/09/938,275
CURRENT FILING DATE: 2001-08-16
NUMBER OF SEQ ID NOS: 11
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 3075
TYPE: PRP
ORGANISM: Homo Sapiens
PUBLICATION INFORMATION:
DATABASE ACCESSION NUMBER: Swissprot P25391
DATABASE ENTRY DATE: 1992-05-01

US-09-938-275-5
Query Match 22.8%; Score 51; DB 10; Length 3075;
Best Local Similarity 28.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 5; Mismatches 17; Indels 14; Gaps 2;

QY 5 NYRYVALKSGSV-----NAPMPENGVTENNDDVFMGYTQ 40
1: | : |||| : |||| :
Db 1253 NEPPQVLINGIRIKROYITMDAPAPENGVRQDEVMARENFKYFNSVSE 1302

RESULT 9

US-09-938-275-5
Query Match 22.8%; Score 51; DB 10; Length 3075;
Best Local Similarity 28.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 5; Mismatches 17; Indels 14; Gaps 2;

QY 5 NYRYVALKSGSV-----NAPMPENGVTENNDDVFMGYTQ 40
1: | : |||| : |||| :
Db 1253 NEPPQVLINGIRIKROYITMDAPAPENGVRQDEVMARENFKYFNSVSE 1302

RESULT 9

US-09-938-275-4
; Sequence 4, Application US/09938275
; Patent No. US2002011309A1
; GENERAL INFORMATION:
; APPLICANT: Gerardo Castillo
; APPLICANT: Alan Snow
; TITLE OF INVENTION: Therapeutic and Diagnostic Applications
; TITLE OF INVENTION: of Laminin and Laminin-Derived Protein Fragments
; FILE REFERENCE: PROTEO.P03
; CURRENT APPLICATION NUMBER: US/09/938,275
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3084
; TYPE: PRT
; ORGANISM: Mus Musculus
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Swissprot P19137
; DATABASE ENTRY DATE: 1990-11-01
US-09-938-275-4

Query Match 22.8%; Score 51; DB 10; Length 3084;
Best Local Similarity 28.0%; Pred. No. 1.6e+02;
Matches 14; Conservative 6; Mismatches 16; Indels 14; Gaps 2;

OY 5 NRYVALKSGS-----VNAPMPENGVTENNND-----WVFMGYTQE 40
DB 1260 NRPQVLLIGRARKKRVITMDAPAPENGVRQDYEVQMKERKYNVNSE 1309

RESULT 10
US-09-815-242-5118
; Sequence 518, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5118
; LENGTH: 1317
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5118

Query Match 22.1%; Score 49.5; DB 10; Length 1317;
Best Local Similarity 33.3%; Pred. No. 97;

Matches 12; Conservative 4; Mismatches 9; Indels 11; Gaps 1;
OY 1 LYXKRYVALKSGSVNAPMPENGVTENNNDWFMG 36
DB 1122 LYXKRYRY-----LPFAGRYASQDPLGLG 1146

RESULT 11
US-10-060-509-20
; Sequence 20, Application US/10060509
; Patent No. US20020120954A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; APPLICANT: Choi, Giltso
; APPLICANT: Johnson, Eric
; APPLICANT: Yi, Hankuil
; APPLICANT: Shin, Byongchul
; TITLE OF INVENTION: Genetic Sequences Encoding Substrate-Specific
; TITLE OF INVENTION: Dihydroflavanol 4-Reductase
; FILE REFERENCE: 4469-111 US
; CURRENT APPLICATION NUMBER: US/10/060,509
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: 09/638,715
; PRIOR FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-10-060-509-20

Query Match 21.9%; Score 49; DB 12; Length 60;
Best Local Similarity 34.6%; Pred. No. 3.4;
Matches 9; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

OY 7 RYVALKSGSVNAPMPENGVTENNNDW 32
DB 7 RFVFTSSAGTVNVEHQKNVYDENDW 32

RESULT 12
US-10-060-506-20
; Sequence 20, Application US/10060506
; Patent No. US20020120959A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; APPLICANT: Choi, Giltso
; APPLICANT: Johnson, Eric
; APPLICANT: Yi, Hankuil
; APPLICANT: Shin, Byongchul
; TITLE OF INVENTION: Genetic Sequences Encoding Substrate-Specific Dihydroflavanol
; FILE REFERENCE: 4469-111 US
; CURRENT APPLICATION NUMBER: US/10/060,506
; CURRENT FILING DATE: 2002-01-30
; PRIOR APPLICATION NUMBER: 09/638,715
; PRIOR FILING DATE: 2000-08-14
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-10-060-506-20

Query Match 21.9%; Score 49; DB 12; Length 60;
Best Local Similarity 34.6%; Pred. No. 3.4;
Matches 9; Conservative 4; Mismatches 13; Indels 0; Gaps 0;

OY 7 RYVALKSGSVNAPMPENGVTENNNDW 32
DB 7 RFVFTSSAGTVNVEHQKNVYDENDW 32

Db 7 REVFTSSAGTVNVEHOKNVDENDW 32

RESULT 13

US-09-846-808-8
 ; Sequence 8, Application US/09846808
 ; Patent No. US20020064531A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Walker, David H.
 ; APPLICANT: Yu, Xu-Tie
 ; TITLE OF INVENTION: Ehrlichia chaffeensis 28 kDa Outer Membrane
 ; FILE REFERENCE: D6311
 ; CURRENT APPLICATION NUMBER: US/09/846,808
 ; CURRENT FILING DATE: 2001-05-01
 ; PRIOR APPLICATION NUMBER: 60/201,035
 ; PRIOR FILING DATE: 2000-05-01
 ; NUMBER OF SEQ ID NOS: 53
 ; SEQ ID NO 8
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Ehrlichia chaffeensis
 ; FEATURE:
 ; OTHER INFORMATION: P28-8 Outer Membrane Protein of
 ; OTHER INFORMATION: Ehrlichia chaffeensis
 US-09-846-808-8

Query Match 21.9%; Score 49; DB 10; Length 275;
 Best Local Similarity 39.3%; Pred. No. 19;
 Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 6 YRYALKSGSVNAPMPENGVTENNNDV 33
 |||:| | | :| | || |
 Db 136 YRFFALARNPSSGSSPTSNNTYVNRNDGV 163

RESULT 14

US-10-059-964-28
 ; Sequence 28, Application US/10059964
 ; Patent No. US20020120115A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Rikihisa, Yasuko
 ; APPLICANT: Ohasi, No. US20020120115A1
 ; TITLE OF INVENTION: Outer Membrane Protein of Ehrlichia Canis and Ehrlichia
 ; TITLE OF INVENTION: Chaffeensis
 ; FILE REFERENCE: 22727/04021
 ; CURRENT APPLICATION NUMBER: US/10/059,964
 ; CURRENT FILING DATE: 2002-01-28
 ; EARLIER APPLICATION NUMBER: 09/314,701
 ; EARLIER FILING DATE: 1999-05-19
 ; NUMBER OF SEQ ID NOS: 66
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 28
 ; LENGTH: 275
 ; TYPE: PRT
 ; ORGANISM: Ehrlichia chaffeensis
 US-10-059-964-28

Query Match 21.9%; Score 49; DB 12; Length 275;
 Best Local Similarity 39.3%; Pred. No. 19;
 Matches 11; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 6 YRYALKSGSVNAPMPENGVTENNNDV 33
 |||:| | | :| | || |
 Db 136 YRFFALARNPSSGSSPTSNNTYVNRNDGV 163

RESULT 15

US-09-960-472-1
 ; Sequence 1, Application US/09960472
 ; Patent No. US20020136712A1
 ; GENERAL INFORMATION:
 ; APPLICANT: FISCHETTI, Vincent

; APPLICANT: LOEFFLER, Julia
 ; APPLICANT: NELSON, Daniel
 ; APPLICANT: LOOMIS, Lawrence
 ; TITLE OF INVENTION: THE USE OF BACTERIAL PHAGE ASSOCIATED LYSING ENZYMES FOR THE P
 ; TITLE OF INVENTION: AND THERAPEUTIC TREATMENT OF COLONIZATION AND INFECTIONS CAUS
 ; FILE REFERENCE: STREPSO-1
 ; CURRENT APPLICATION NUMBER: US/09/960,472
 ; CURRENT FILING DATE: 2001-09-21
 ; PRIOR APPLICATION NUMBER: US 09/846,688
 ; PRIOR FILING DATE: 2001-05-02
 ; PRIOR APPLICATION NUMBER: US 09/497,495
 ; PRIOR FILING DATE: 2000-04-18
 ; PRIOR APPLICATION NUMBER: US 09/395,636
 ; PRIOR FILING DATE: 1999-09-14
 ; PRIOR APPLICATION NUMBER: US 08/962,523
 ; PRIOR FILING DATE: 1997-10-31
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 1
 ; LENGTH: 296
 ; TYPE: PRT
 ; ORGANISM: bacteriophage Dp-1
 US-09-960-472-1

Query Match 21.9%; Score 49; DB 10; Length 296;
 Best Local Similarity 33.3%; Pred. No. 21;
 Matches 13; Conservative 7; Mismatches 5; Indels 14; Gaps 2;

QY 8 YVALKSGS-----VNAPMPENG---VTENNNDW 32
 |||:| | | :| | || |
 Db 39 YVALRSAGASAGAVNTEYHMAWLIENGYELISENAPW 77

Search completed: December 10, 2002, 10:58:07
 Job time : 6.2 secs

THIS PAGE BLANK (USPTO)


```
; CURRENT APPLICATION NUMBER: PCT/US02/35143
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/335,371
; PRIOR FILING DATE: 2001-11-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 417
; TYPE: PR1
; ORGANISM: Rhodococcus
PCT-US02-35143-14
```

```
Query Match          23.2%; Score 52; DB 1; Length 417;
Best Local Similarity 36.7%; Pred. No. 23;
Matches 11; Conservative 6; Mismatches 11; Indels 2; Gaps 1;
```

```
QY 12 KSGGSVNPMPEN--GVTENNNDVFMGYTQ 39
      | : | : | : | : | : | : | : | : |
Db 183 QQGAIIAAIPTSRAGVTPMDMAIGMRQ 212
```

```
RESULT 3
US-10-286-326-14
; Sequence 14, Application US/10286326
; GENERAL INFORMATION:
; APPLICANT: San, Ka-Yui
; APPLICANT: Berrios-Rivera, Susana
; APPLICANT: Bennett, George
; TITLE OF INVENTION: Recycling System for Manipulation of Intracellular NADH Availabil
; FILE REFERENCE: P02328051
; CURRENT APPLICATION NUMBER: US/10/286,326
; CURRENT FILING DATE: 2002-11-01
; PRIOR APPLICATION NUMBER: US 60/335,371
; PRIOR FILING DATE: 2001-11-02
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 417
; TYPE: PR1
; ORGANISM: Rhodococcus
US-10-286-326-14
```

```
Query Match          23.2%; Score 52; DB 6; Length 417;
Best Local Similarity 36.7%; Pred. No. 23;
Matches 11; Conservative 6; Mismatches 11; Indels 2; Gaps 1;
```

```
QY 12 KSGGSVNPMPEN--GVTENNNDVFMGYTQ 39
      | : | : | : | : | : | : | : | : |
Db 183 QQGAIIAAIPTSRAGVTPMDMAIGMRQ 212
```

```
RESULT 4
US-09-134-000C-3435
; Sequence 3435, Application US/09134000C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3435
; LENGTH: 167
; TYPE: PR1
; ORGANISM: Enterococcus faecalis
US-09-134-000C-3435
```

```
Query Match          22.8%; Score 51; DB 5; Length 167;
Best Local Similarity 32.1%; Pred. No. 12;
```

```
Matches 9; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
QY 1 LXXKNRYVALKSGGSVNPMPENGYTE 28
      | : | : | : | : | : | : | : | : |
Db 83 IYWMKRNHTSAFSGTDIDIRLRERGITTE 110
```

```
RESULT 5
PCT-US02-34452-4
; Sequence 4, Application PC/YUS0234452
; GENERAL INFORMATION:
; APPLICANT: INCYTE GENOMICS, INC.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: ELLIOTT, Vicki S.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: LI, Joana X.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: DELEGANE, Angelo M.
; APPLICANT: LEE, Soo Yeun
; APPLICANT: BECHA, Shanya D.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: KABLE, Amy E.
; APPLICANT: MARQUIS, Joseph P.
; APPLICANT: GURURAJAN, Rajagopal
; APPLICANT: SPRAGUE, William W.
; APPLICANT: YANG, Junming
; APPLICANT: GIEZEN, Kimberly J.
; APPLICANT: ZEBARADJIAN, Yeganeh
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: JACKSON, Alan A.
; APPLICANT: JIANG, Xin
; TITLE OF INVENTION: VESICLE-ASSOCIATED PROTEINS
; FILE REFERENCE: PF-1234 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/34452
; CURRENT FILING DATE: 2002-10-24
; PRIOR APPLICATION NUMBER: US 60/347,927
; PRIOR FILING DATE: 2001-10-26
; PRIOR APPLICATION NUMBER: US 60/332,908
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/331,865
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/342,604
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: 60/354,827
; PRIOR FILING DATE: 2002-02-06
; NUMBER OF SEQ ID NOS: 40
; SOFTWARE: PERL Program
; SEQ ID NO 4
; LENGTH: 648
; TYPE: PR1
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 7503143CD1
PCT-US02-34452-4
```

```
Query Match          22.8%; Score 51; DB 1; Length 648;
Best Local Similarity 48.1%; Pred. No. 52;
Matches 13; Conservative 3; Mismatches 7; Indels 4; Gaps 2;
```

```
QY 9 YALKSGSVNPMPENGVTENN--DW 32
      | : | : | : | : | : | : | : | : |
Db 373 YAIWTGSDV-APMGREGVTAMHKLFDW 398
```

```
RESULT 6
US-09-724-676-72290
; Sequence 72290, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
```

```
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: Patentin version 3.2
SEQ ID NO 72290
LENGTH: 2606
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676-72290

Query Match      22.8%; Score 51; DB 5; Length 2606;
Best Local Similarity 28.0%; Pred. No. 2.4e+02;
Matches 14; Conservative 5; Mismatches 17; Indels 14; Gaps 2;

Y 5 NRYRYALKSGGSV-----NAPPENGVTE-----NNDWVFMGYTQE 40
D 1253 NPEQVLIKGRIRKQYIYMDAPAPENGVRQEOEVAMRENFMYFNSVSE 1302

RESULT 7
US-09-724-676A-72290
Sequence 72290: Application US/09724676A
GENERAL INFORMATION:
APPLICANT: Compugen LTD
TITLE OF INVENTION: Variants of alternative splicing
FILE REFERENCE: 129181.4 Compugen
CURRENT APPLICATION NUMBER: US/09/724,676A
CURRENT FILING DATE: 2000-11-28
NUMBER OF SEQ ID NOS: 97222
SOFTWARE: Patentin version 3.2
SEQ ID NO 72290
LENGTH: 2606
TYPE: PRT
ORGANISM: Homo sapiens
US-09-724-676A-72290

Query Match      22.8%; Score 51; DB 5; Length 2606;
Best Local Similarity 28.0%; Pred. No. 2.4e+02;
Matches 14; Conservative 5; Mismatches 17; Indels 14; Gaps 2;

Y 5 NRYRYALKSGGSV-----NAPPENGVTE-----NNDWVFMGYTQE 40
D 1253 NPEQVLIKGRIRKQYIYMDAPAPENGVRQEOEVAMRENFMYFNSVSE 1302

RESULT 8
US-10-284-060-59
Sequence 59: Application US/10284060
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: French-Constant, Richard
APPLICANT: Orr, Gregory L
APPLICANT: Merlo, Donald J
APPLICANT: Roberts, Jean L
APPLICANT: Rocheleau, Thomas A
TITLE OF INVENTION: Insecticidal Protein Toxins from
TITLE OF INVENTION: Photornabads
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dowelanco
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/284,060
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/063,615
FILING DATE: 18-MAY-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/395,497
FILING DATE: 28-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/007,255
FILING DATE: 06-NOV-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/608,423
FILING DATE: 28-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/705,484
FILING DATE: 28-AUG-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/743,699
FILING DATE: 06-NOV-1996
ATTORNEY/AGENT INFORMATION:
NAME: Borucki, Andrea T.
REGISTRATION NUMBER: 33651
REFERENCE/DOCKET NUMBER: 50301E
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-337-4846
TELEFAX: 317-337-4847
INFORMATION FOR SEQ ID NO: 59:
SEQUENCE CHARACTERISTICS:
LENGTH: 1565 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-10-284-060-59

Query Match      22.3%; Score 50; DB 6; Length 1565;
Best Local Similarity 40.5%; Pred. No. 1.9e+02;
Matches 17; Conservative 2; Mismatches 13; Indels 10; Gaps 3;

Y 1 LVY-----KNRYRYALKSGGSVNAPME-NGVTEN--NDW 32
D 180 IYFICRTTKYRYIWRQMDLSKNRODPAGNPVTPNCWMDW 221

RESULT 9
US-10-284-060-59
Sequence 59: Application US/10284060
GENERAL INFORMATION:
APPLICANT: Ensign, Jerald C
APPLICANT: Bowen, David J
APPLICANT: Petell, James
APPLICANT: Fatig, Raymond
APPLICANT: Schoonover, Sue
APPLICANT: French-Constant, Richard
APPLICANT: Orr, Gregory L
APPLICANT: Merlo, Donald J
APPLICANT: Roberts, Jean L
APPLICANT: Rocheleau, Thomas A
TITLE OF INVENTION: Insecticidal Protein Toxins from
TITLE OF INVENTION: Photornabads
NUMBER OF SEQUENCES: 88
CORRESPONDENCE ADDRESS:
ADDRESS: Dowelanco
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US
ZIP: 46268
COMPUTER READABLE FORM:
```

STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: IN
COUNTRY: US

```

RESULT 11
US-09-134-000C-6557
: Sequence 6557, Application US/09134000C
:
: GENERAL INFORMATION:
: APPLICANT: Lynn Doucette-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
: TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 032796-032
: CURRENT APPLICATION NUMBER: US/09/134, 000C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/055,778
: PRIOR FILING DATE: 1997-08-15
: NUMBER OF SEQ ID NOS: 6812
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 6557
: LENGTH: 439
: TYPE: PR1
: ORGANISM: Enterococcus faecalis
: FEATURE:
: NAME/KEY: MISC_FEATURE
: LOCATION: (439)..(439)

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:50:48 ; Search time 11 Seconds

(without alignments)
349.580 Million cell updates/sec

Title: US-09-142-970-2_COPY_1_40

Perfect score: 224
Sequence: 1 LYKKNRYALKSGSVNAPMPENGVTENNDFWFGYTOE 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	210	93.8	1532	2 A26039	IgA-specific metal
2	202	90.2	1561	2 S61314	IgA-specific metal
3	202	90.2	1773	2 A81937	IgA-specific metal
4	198	88.4	1815	2 C81169	IgA-specific metal
5	198	88.4	1541	2 A37023	IgA-specific metal
6	98	43.8	1694	2 H64106	IgA-specific metal
7	98	43.8	1702	2 A41859	IgA-specific metal
8	90	40.2	1849	2 C41859	IgA-specific metal
9	87	38.8	1545	2 B41859	IgA-specific metal
10	62.5	27.9	532	1 CPBY	carboxypeptidase C
11	58	25.9	297	2 S55085	D-alanine D-alanine
12	58	25.9	452	2 B90640	D-alanine D-alanine
13	58	25.9	452	2 B85491	UDP-N-acetylmutam
14	58	25.9	452	2 A10517	UDP-N-acetylmutam
15	58	25.9	452	2 F64730	UDP-N-acetylmutam
16	56	25.0	359	2 A43532	B-cell surface ant
17	56	25.0	574	2 A91182	hypothetical prote
18	56	25.0	574	2 E86028	hypothetical prote
19	56	25.0	821	2 C64461	hypothetical prote
20	56	25.0	1367	2 T33819	hypothetical prote
21	54	24.1	350	2 S47292	phenol 2-monooxyg
22	54	24.1	577	2 AD1440	hypothetical prote
23	53.5	23.9	648	2 S70907	transferrin-bindin
24	53.5	23.9	658	2 E95111	endo-beta-N-acetyl
25	53.5	23.9	671	2 A38109	autolysin - Entero
26	53.5	23.9	721	2 C97980	endo-beta-N-acetyl
27	53.5	23.9	766	2 T20003	hypothetical prote
28	53.5	23.9	850	2 S20462	RNA12 protein - ye
29	52.5	23.4	1577	2 T30858	glucosyltransferas

My teen

ALIGNMENTS

30	52	23.2	221	2 D90603	hypothetical prote
31	52	23.2	478	2 F90497	hypothetical prote
32	52	23.2	515	2 S20493	endoglycanase - Cl
33	52	23.2	598	2 T33240	hypothetical prote
34	52	23.2	1356	2 T16718	hypothetical prote
35	51.5	23.0	114	2 AD0785	conserved hypothet
36	51.5	23.0	213	2 B86828	hypothetical prote
37	51.5	23.0	293	2 C75421	hypothetical prote
38	51.5	23.0	306	2 G88711	protein C17H12.8 l
39	51.5	23.0	362	2 A12113	hypothetical prote
40	51.5	23.0	682	2 G86300	F19K19.13 protein
41	51.5	23.0	705	2 S51396	glycogen(starch) s
42	51.5	23.0	993	2 A38437	probable homeotic
43	51	22.8	106	2 E88188	protein C18H9.6 l
44	51	22.8	291	2 S03230	hypothetical prote
45	51	22.8	574	2 D65153	64.9 kD protein in

RESULT 1

A26039
IgA-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (s
N:Alternate names: IgA protease; immunoglobulin A1 proteinase
C:Species: Neisseria gonorrhoeae
A:Variety: strain MS11
C:Date: 05-Oct-1988 #sequence-revision 05-Oct-1988 #text-change 08-Dec-2000
C:Accession: A26039; S09386
R:Pohlner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 48-462, 1987
A:Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA prot
A:Reference number: A26039; PMID:87115823; PMID:3027577
A:Accession: A26039
A:Molecule type: DNA
A:Residues: 11532 <POH>
A:Cross-references: GB:X04835; NID:944868; PIDN:CAA28538.1; PID:944869
A:Note: The authors translated the codon AAG for residue 668 as Asn
R:Halter, R.; Pohlner, J.; Meyer, T.F.
EMBO J 8, 2737-2744, 1989
A:Title: Mosaic-like organization of IgA protease genes in Neisseria gonorrhoeae gene
A:Reference number: S09386; PMID:9060056; PMID:2511009
A:Accession: S09386
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'W', 429-531, 'N', 533-615, 'V', 617-631, 'N', 6
A:Experimental source: strain MS11
C:Genetics:
A:Gene: Iga
C:Superfamily: IgA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <Stc>
F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <Mat>
F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match
Best Local Similarity 93.8%; Score 210; DB 2; Length 1532;
Pred. No. 4.9e-19;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

DB 584 LYKKNRYALKSGSVNAPMPENGVTENNDFWFGYTOE 623

RESULT 2
S61314
IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: Hfl3
C:Date: 20-Jul-1996 #sequence-revision 13-Mar-1997 #text-change 08-Dec-2000
C:Accession: S61314

R:lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1561 <LOW>
A:Cross-references: EMBL:X82474; NID:g732873; PIND:CAAS7857.1; PID:g732874
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 90.2%; Score 202; DB 2; Length 1561;
Best Local Similarity 90.0%; Pred. No. 5,16e-18;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 40
DB 584 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 623

RESULT 3

Iga-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - Neisseria meningitidis
N:Alternate names: Iga1 proteinase; Iga1 proteinase (EC 3.4.21.7) [misnomer]; Immunoglobulin
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain HF117; strain HF159; strain SM1027
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtz, S.; Jørgensen, K.; Leather, S.; Mouton, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIND:CA884182.1; PID:g737961
A:Experimental source: serogroup A, strain 22491
R:lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neisseria
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82470; NID:g732854; PIND:CAAS7853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82471; NID:g732858; PIND:CAAS7854.1; PID:g732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82472; NID:g732852; PIND:CAAS7855.1; PID:g732853
A:Experimental source: strain SM1027
C:Genetics:
A:Gene: iga; NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 90.2%; Score 202; DB 2; Length 1773;
Best Local Similarity 90.0%; Pred. No. 6,4e-18;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 40
DB 573 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 612

RESULT 4

Iga-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: strain MC58 serogroup B; strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Petelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eissen,
Hickey, E.K.; Haft, D.H.; Salzman, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
rl, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:20157555; PMID:10710307
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <DET>
A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIND:AAFA1117.1; PID:g722
A:Experimental source: serogroup B, strain MC58
R:lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82477; NID:g732856; PIND:CAAS7860.1; PID:g732857
C:Genetics:
A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 88.4%; Score 198; DB 2; Length 1815;
Best Local Similarity 87.5%; Pred. No. 2,2e-17;
Matches 35; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 40
DB 573 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 612

RESULT 5

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influenzae
N:Alternate names: Immunoglobulin A1 proteinase type 1
C:Species: Haemophilus influenzae
C:Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C:Accession: A37023
R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Killian, M.
Infect. Immun. 57, 3097-3105, 1989
A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of Haemo
A:Reference number: A37023; MUID:89379374; PMID:2506130
A:Accession: A37023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1541 <POU>
A:Cross-references: GB:X64357; NID:g43560; PIND:CAAS7808.1; PID:g43561
A:Experimental source: serotype b
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 43.8%; Score 98; DB 2; Length 1541;
Best Local Similarity 40.0%; Pred. No. 0,00022;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNDFWFMGYTOE 40
DB 595 LNNENYTYALRKGASTRSELPKNSGSENNWLMKGTSD 634

RESULT 6

H64106
IGA-specific metalloendopeptidase (EC 3.4.24.13) type I - Haemophilus influenzae (strain N:Altemate names: immunoglobulin A1 proteinase type 1
C:Species: Haemophilus influenzae
A:Variety: strain Rd KW20
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000
C:Accession: H64106; A41500
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Gacyaynchman, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fitchman, J.L.; Fuhrmann, J.L.; Georgagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Grehn, C.L., McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.
A>Title: Whole genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MOID:95350630; PMID:7542800
A:Accession: H64106
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1694 <TIG>
A:Cross-references: GB:U32779; GB:I42023; NID:g1574009; PIDN:AAC22651.1; PID:g1574019; T
A:Experimental source: Strain Rd KW20
R:Grund, F.J.; Plaut, A.G.; Wright, A.
Infect. Immun. 58, 320-331, 1990
A>Title: Localization of the cleavage site specificity determinant of Haemophilus influe
A:Reference number: A41500; MOID:90129281; PMID:2105270
A:Accession: A41500
A:Molecule type: DNA
A:Residues: 1-377 <GRU>
A:Cross-references: GB:X59800
A:Experimental source: strain Rd KW20
A>Note: The authors translated the codon TGG for residue 319 as Thr
C:Function:
A:Description: this proteinase is classified as type 1 because it cleaves at a proline-s
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 43.8%; Score 98; DB 2; Length 1694;
Best Local Similarity 40.0%; Pred. No. 0.00024;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGSYNAPMPENGVTENNNDWFMGTQOE 40
| :|| ||||: | :| :||: | :||:| | :
DB 601 LNEENTTYALRKNGASTRSELPKNSGESNNMWTMGKTSD 640

RESULT 7

A41859
IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influen
C:Species: Haemophilus Influenzae
A:Variety: strain HK715
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000
C:Accession: A41859
R:Poulsen, K.; Reinholdt, J.; Kilian, M.
J. Bacteriol. 174, 2913-2921, 1992
A>Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty
A:Reference number: A41859; MOID:92234949; PMID:1373717
A:Accession: A41859
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1702 <POD>
A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907
A:Experimental source: strain HK715
A>Note: sequence extracted from NCBI backbone (NCBIP:97282)
C:Superfamily: IGA-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 43.8%; Score 98; DB 2; Length 1702;
Best Local Similarity 40.0%; Pred. No. 0.00024;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGSYNAPMPENGVTENNNDWFMGTQOE 40
| :|| ||||: | :| :||: | :||:| | :
DB 601 LNEENTTYALRKNGASTRSELPKNSGESNNMWTMGKTSD 640

[illegible]

A:Molecule type: DNA
A:Residues: 1-532 <VAL>
A:Cross-references: EMBL:MI5482; NID:g172238; PIDN:AAA34902.1; PID:g172239
R:Barrell, B.G.
submitted to the EMBL Data Library, August 1994
A:Reference number: S47445
A:Accession: S47458
A:Molecule type: DNA
A:Residues: 1-532 <BAR>
A:Cross-references: EMBL:X80836; NID:g1289327; PIDN:CAA56806.1; PID:g530354; GSPDB:GN0004
R:Svendsen, I.; Martin, B.M.; Vismannatha, T.; Johansen, J.T.
Carlsberg Res. Commun. 47, 15-27, 1982
A:Title: Amino acid sequence of carboxypeptidase Y. II. Peptides from enzymatic cleavage
A:Reference number: A90763
A:Accession: A90763
A:Molecule type: Protein
A:Residues: 112-223, 'X', 225, 228-239, 'X', 241-259, 'HG', 262-267, 'X', 269-388, 'E', 390-451, 'N'
R:Svendsen, I.
submitted to the Atlas, October 1982
A:Reference number: A94609
A:Contents: disulfide bond
A:Accession: A94609
A:Molecule type: protein
A:Residues: 224-227 <SV2>
A>Note: this is a revision to the sequence in reference A90763
R:Martin, B.M.; Svendsen, I.; Vismannatha, T.; Johansen, J.T.
Carlsberg Res. Commun. 47, 1-13, 1982
A:Title: Amino acid sequence of carboxypeptidase Y. Peptides from cleavage with cyanogen
A:Reference number: A90762
A:Contents: annotation; experimental details
R:Endrizzi, J.A.; Remington, S.J.
submitted to the Brookhaven Protein Data Bank, March 1994
A:Reference number: A52472; PDB:1VSC
A:Contents: annotation; x-ray crystallography at 2.8 angstroms; disulfide bonds
C:Genetics:
A:Gene: SGD:PRC1; lbc1; MIPS:YMR297w
A:Cross-references: SGD:S0004912; MIPS:YMR297w
A:Map position: 13R
C:Superfamily: serine carboxypeptidase
C:Keywords: glycoprotein; hydrolase; serine carboxypeptidase; yeast vacuole; zymogen
F:1-50/Domain: signal sequence #status predicted <SIG>
F:21-111/Domain: propeptide #status predicted <PRO>
F:112-532/Product: carboxypeptidase C #status experimental <MANT>
F:124,188,279/Binding site: carbonylhydride (Asn) (covalent) #status predicted
F:167-409,304-318,328-351,335-344,373-379/Disulfide bonds: #status experimental
F:257/Active site: Ser #status experimental
F:449,508/Active site: Asp, His #status predicted

Query Match	27.9%	Score 62.5;	DB 1;	Length 532;
Best Local Similarity	32.4%	Pred. No. 3;		
Matches 12:	Conservatively 11;	Mismatches 13;	Indels 1;	Gaps 1;

OY 3 YKNRYALKSGG-SVNAPMPENGTEENNDFWEMQYT 38
|::||::||::||::||::||::||::||::||::||::||
Db 495 YKHFTYLAVFNCGHNVPPDVEPNALSMYNEMIHGSFS 531

RESULT 11
S55085
hypothetical protein YMR099c - yeast (*Saccharomyces cerevisiae*)
N:Alternate names: hypothetical protein YM6543.06c
C:Species: *Saccharomyces cerevisiae*
C>Date: 08-Jul-1995 #sequence.revision 01-Sep-1995 #text_change 19-Apr-2002
C:Accession: S55085
R:Hunt, S.; Bowman, S.
submitted to the EMBL Data Library, June 1995
A:Reference number: S55080
A:Accession: S55085
A:Molecule type: DNA
A:Residues: 1-297 <HUN>
A:Cross-references: EMBL:Z49807; NID:g854430; PIDN:CAA89900.1; PID:g854436; GSPDB:GN0001
I:Experimental source: strain AB972

```

C:Genetics:
A:Gene: MIPS:YMR099c
A:Cross-references: SGD:S0004705
A:Map position: 13R
C:Superfamily: Arabidopsis thaliana hypothetical protein F9D16.200

Query Match          25.9%   Score 58; DB 2; Length 297;
Best Local Similarity 40.0%; Pred. No. 6;
Matches      8; Conservative    7; Mismatches     5; Indels     0; Gaps     0;

QY      21 MPENGVTENNNDWVFMGYTQE 40
       ::::|::|::|::|::|::|
Db      79 LPQHGLARNSTWEFLGQTKE 98

RESULT 12
B90640
D-alanine D-alanine-adding enzyme [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90640
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.
gasakara, N.; Yasunaga, T.; Kunihara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90640
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-452 <HAY>
A:Cross-references: GB:BA000007; PIDN:BAH3513.1; PID:g13359546; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: ECS0090

Query Match          25.9%   Score 58; DB 2; Length 452;
Best Local Similarity 36.3%; Pred. NO. 9.6;
Matches      10; Conservative    7; Mismatches     5; Indels     4; Gaps     1;

QY      12 KSGGSVNAPMPCNV---TENNDMV 33
       |::|::|::|::|::|::|
Db      200 KAKGEIFSGLPENGAIMNADNDML 225

RESULT 13
B85491
D-alanine D-alanine adding enzyme [imported] - Escherichia coli (strain O157:H7, subs
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85491
R:Penna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
liller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamoustis, K.; Apoda
Native 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: B85491
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-452 <STO>
A:Cross-references: GB:AE005174; NID:g12512789; PIDN:AAG54390.1; GSPDB:GN00145; UWGP:
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: mufF

Query Match          25.9%   Score 58; DB 2; Length 452;
Best Local Similarity 38.5%; Pred. No. 9.6;
Matches      10; Conservative    7; Mismatches     5; Indels     4; Gaps     1;

QY      12 KSGGSVNAPMPCNV---TENNDMV 33
       |::|::|::|::|::|::|
Db      200 KAKGEIFSGLPENGAIMNADNDML 225

```


THIS PAGE BLANK (USPTO)


```

KW Transmembrane: Signal.
FT SIGNAL 1 27
FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.
FT PROPEP 987 1532 HELPER PEPTIDE.
FT ACT SITE 278 278 POTENTIAL.
FT SITE 986 987 CLEAVAGE (AUTO-).
FT SITE 1018 1019 CLEAVAGE (AUTO-).
FT SITE 1121 1122 CLEAVAGE (AUTO-).
SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD2F40D CRC64;

Query Match
Best Local Similarity 93.8%; Score 210; DB 1; Length 1532;
Matches 36; Conservative 2; Mismatches 20; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGSVNAPMPENGVTENNDDVFMGTQOE 40
DB 584 LYKKNRYALKSGGRNLNAPMPENGVAENNDDIMFGTQOE 623

RESULT 2
IGA1_HAEIN STANDARD; PRT; 1541 AA.
ID IGA1_HAEIN
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OY NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK368 / Serotype B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RL of Haemophilus influenzae serotype b.";
RN Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SFR-288.
RC STRAIN-HK368 / Serotype B;
RX MEDLINE=92234949; PubMed=137317;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RL influenzae type 1 immunoglobulin A1 proteases.";
RN J. Bacteriol. 174:2913-2921(1992).
CC -!- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: X64357; CAA45708.1;
CC DR EMBL: M87492; AAA24969.1;
CC DR MEROPS: S06.001;
CC InterPro: IPR000710; IGA_S6.

```

```

DR InterPro: IPR004899; Pectact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF02312; Pectactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008 POTENTIAL.
FT PROPEP 1009 1541 IMMUNOGLOBULIN A1 PROTEASE.
FT ACT SITE 288 288 HELPER PEPTIDE (POTENTIAL).
FT MUTAGEN 288 288 PROBABLE.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

Query Match
Best Local Similarity 43.8%; Score 98; DB 1; Length 1541;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGSVNAPMPENGVTENNDDVFMGTQOE 40
DB 595 LNLNENTYALRRKGASTRELPRKSGSENNMLYMGKTSQ 634

RESULT 3
IGA0_HAEIN STANDARD; PRT; 1694 AA.
ID IGA0_HAEIN
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR H10990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OY NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL submitted (May-1991) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Kennedy K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glöck A., Kelley J.M.,
RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ullrich T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd.";
RN Science 269:496-512(1995).
CC -!- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
CC EMBL: X59800; ? NOT ANNOTATED_CDS.
CC EMBL: U32779; AAC22651.1; ?
CC MEROPS: S06.001; ?
CC DR TIGR: H10990; ?
CC DR InterPro: IPR000710; IGA_S6.
CC DR InterPro: IPR004899; Pertact_sup.
CC Pfam: PF02395; IGA1; 1.
CC DR Pfam: PF03212; Pertactin; 2.
CC DR PRINTS: PR00921; IGASERTPASE.
CC KM Hydrolyase; Serine protease; Transmembrane; Zymogen; signal;
CC Complete proteome.
CC FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> A (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> T (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;

Query Match 43.8%; Score 98; DB 1; Length 1694;
Best Local Similarity 40.0%; Pred. No. 7.1e-05;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGSVNAPMPENGVTENNDFMGYTOE 40
DB 601 LNLNNTYVALRKGASTRSELPKSGSENNMLYMGKTS D 640

RESULT 4
IGA2_HAEIN STANDARD: PRT: 1702 AA.
ID IGA2_HAEIN
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RT J. Bacteriol. 174:2913-2921(1992).
RU J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
certain Pro-I-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-I-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC -----
CC EMBL: M87489; AAA24966.1; ?
CC MEROPS: S06.001; ?
CC DR InterPro: IPR000710; IGA_S6.
CC DR InterPro: IPR004899; Pertact_sup.
CC Pfam: PF02395; IGA1; 1.
CC DR Pfam: PF03212; Pertactin; 2.
CC DR PRINTS: PR00921; IGASERTPASE.
CC KM Hydrolyase; Serine protease; Transmembrane; Zymogen; Repeat; signal.
CC Complete proteome.
CC FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 x 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
K.
FT REPEAT 1109 1116 1.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F7002667807A6 CRC64;

Query Match 43.8%; Score 98; DB 1; Length 1702;
Best Local Similarity 40.0%; Pred. No. 7.2e-05;
Matches 16; Conservative 9; Mismatches 15; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGSVNAPMPENGVTENNDFMGYTOE 40
DB 601 LNLNNTYVALRKGASTRSELPKSGSENNMLYMGKTS D 640

RESULT 5
IGA4_HAEIN STANDARD: PRT: 1849 AA.
ID IGA4_HAEIN
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1 HK61;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
certain Pro-I-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-I-xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

```

CC or send an email to license@isb.ch).
CC -----
DR EMBL: M87491; AAA24968.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin-sup.
DR Pfam: PF03395; IGA1; 1.
DR PRINTS: PR00921; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021
FT PROPEP 1022 1849
FT ACT_SITE 299 299
FT ACT_SITE 299 299
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match
Best Local Similarity 40.2%; Score 90; DB 1; Length 1849;
Matches 16; Conservative 8; Mismatches 16; Indels 2; Gaps 1;

QY 1 LYYK--NRYRYALKSGSYNAPMPENGVTENNDDWYEMGYTQE 40
DB 604 LYFNODNRSTYTLKKKASTRSELPQNSGSENNWLMGMTSD 645

RESULT 6
ID IGA3_HAEIN STANDARD; PRT; 1545 AA.
AC P45385;
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulson K., Reinhold J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Asa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE. AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.stb.ch).
CC -----
DR EMBL: M87490; AAA24967.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin-sup.
DR Pfam: PF03395; IGA1; 1.
DR Pfam: PF03395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.

```

```

DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1012
FT PROPEP 1013 1545
FT ACT_SITE 292 292
FT ACT_SITE 292 292
SQ SEQUENCE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;

Query Match
Best Local Similarity 38.8%; Score 87; DB 1; Length 1545;
Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 1;

QY 1 LYY--KNRYRYALKSGSYNAPMPENGVTENNDDWYEMG 36
DB 597 LYFNEENRTYTLKKKASTRSELPQNSGSENNWLMGMTSD 634

RESULT 7
ID CBPY_YEAST STANDARD; PRT; 532 AA.
AC P00729;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
DE PRC1 OR YMR297W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131100; PubMed=3028649;
RA Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;
RT "Protein sorting in yeast: the localization determinant of yeast
RT vacuolar carboxypeptidase Y resides in the propeptide."
RT Cell 48:887-897(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288c / AB972;
RA Baddock K., Churcher C., Barrell B.G., Rajndream M.A.;
RL Submitted (Aug-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 112-532.
RA Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;
RT "Amino acid sequence of carboxypeptidase Y. II. Peptides from
RT enzymatic cleavages."
RT Carlsberg Res. Commun. 47:15-27(1982).
RN [4]
RP REVISIONS, AND ACTIVE SITE SER-257.
RA Bredam K., Svendsen I.;
RT "Identification of methionyl and cysteinyl residues in the substrate
RT binding site of carboxypeptidase Y."
RT Carlsberg Res. Commun. 49:639-645(1984).
RN [5]
RP ACTIVE SITE HIS-508.
RX MEDLINE=90315013; PubMed=2639680;
RA Bech L.M., Bredam K.;
RT "Inactivation of carboxypeptidase Y by mutational removal of the
RT putative essential histidyl residue."
RT Carlsberg Res. Commun. 54:165-171(1989).
RN [6]
RP MUTAGENESIS.
RX MEDLINE=94114535; PubMed=7904479;
RA Mortensen U.H., Remington S.J., Bredam K.;
RT "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen
RT bond network stabilizes the transition state by interaction with the
RT C-terminal carboxylate group of the substrate."
RL Biochemistry 33:508-517(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=95244421; PubMed=7727362;
RA Endrizzi J.A., Bredam K., Remington S.J.;

```

RT "2.8-A structure of yeast serine carboxypeptidase.";
 RL Biochemistry 33:1106-1120(1994).
 CC -I- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS
 CC PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC
 CC RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR
 CC PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
 CC -I- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -I- ENZYME REGULATION: INHIBITED BY 2PCK.
 CC -I- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
 CC -I- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZIMOGEN AND
 CC IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO
 CC A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE
 CC GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE
 CC GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
 CC SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS
 CC SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR
 CC ENDOSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
 CC N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY
 CC ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 10.
 CC -I- DATABASE: NAME=worthington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COY.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: M15482; AAA34902.1; -
 DR EMBL: X80836; CAA56806.1; -
 DR PIR: A26597; CPBY..
 DR PDB: 1YSC; 22-JUN-94.
 DR PDB: 1CPY; 15-SEP-95.
 DR MEROPS: S10.001; -
 DR SGD: S0004912; PRCL.
 DR InterPro: IPR000379; Ser. esters. site.
 DR InterPro: IPR001563; Serine, carboxypept.
 DR Pfam: PF00450; Serine, carboxypept. 1.
 DR PRINTS: PR00724; CRBOXYPTASC.
 DR ProDom: PD001189; Serine, carboxypept. 1.
 DR PROSITE: PS00131; CARBOXYPEPT. SER. SER. 1.
 DR PROSITE: PS00560; CARBOXYPEPT. SER. HIS. 1.
 DR Hydrolyse: Carboxypeptidase; Glycoprotein; Zymogen; Signal;
 DR 3D-structure.
 KM
 FT SIGNAL 1 20
 FT PROPEP 21 111
 FT CHAIN 112 532
 FT ACT_SITE 257 257
 FT ACT_SITE 449 449
 FT ACT_SITE 508 508
 FT BINDING 452 452
 FT BINDING 509 509
 FT DISULFID 167 409
 FT DISULFID 304 318
 FT DISULFID 328 351
 FT DISULFID 335 344
 FT DISULFID 373 379
 FT CARBOHYD 124 124
 FT CARBOHYD 198 198
 FT CARBOHYD 279 279
 FT CARBOHYD 479 479
 FT SITE 24 27
 FT MUTAGEN 508 508
 FT CONFLICT 260 261
 FT CONFLICT 389 389
 FT CONFLICT 529 529
 FT SEQUENCE 532 AA; 59802 MW; 7227F3489CBB952 CRC64;
 Query Match 27.9%; Score 62.5; DB 1; Length 532;

Best Local Similarity 32.4%; Pred. No. 1;
 Matches 12; Conservative 11; Mismatches 13; Indels 1; Gaps 1;
 OY 3 YKMYRYALKSG-SYNAPPEPENGVTENNDDVFMGT 38
 DB 495 YKHFTYLRVFGHMPVFPDENALSMVMEWINGFS 531
 RESULT 8
 YMR9_YEAST STANDARD; PRT; 297 AA.
 ID YMR9_YEAST
 AC Q03161;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 34.0 kDa protein in CTF13-YRK2 intergenic region.
 GN YMR099C OR YMR643.06C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (JUN-1995) to the EMBL/GenBank/DDI databases.
 CC -I- SIMILARITY: BELONGS TO THE UPF010 FAMILY. SIGNIFICANTLY RELATED
 CC TO MURINOTASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 249807; CAA89900.1; -
 DR SGD: S0004705; YMR099C.
 DR InterPro: IPR001823; Ald1-epimerase.
 DR Pfam: PF01263; Aldose-epim. 1.
 DR Hypothetical protein.
 FT ACT_SITE 159 159
 FT SEQUENCE 297 AA; 33956 MW; 122BDC460B9A32B CRC64;
 Query Match 25.9%; Score 58; DB 1; Length 297;
 Best Local Similarity 40.0%; Pred. No. 2.1;
 Matches 8; Conservative 7; Mismatches 5; Indels 0; Gaps 0;
 OY 21 MPENGVTENNDDVFMGTQE 40
 DB 79 LPDHGLARNSTWFFLGOTKE 98
 RESULT 9
 MURF_ECOLI STANDARD; PRT; 452 AA.
 ID MURF_ECOLI
 AC P11880; P77636; 007100;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE UDP-N-acetylmuuramylalanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
 DE alanyl ligase (EC 6.3.2.15) (UDP-murNac-pentapeptide synthetase)
 DE (D-alanyl-D-alanine-adding enzyme).
 GN MURF OR MRA OR B0086.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=89345095; PubMed=2668880;

RA Parquet C., Flouret B., Mengin-Lecreulx D., van Heijenoort J.;
 RT "Nucleotide sequence of the murf gene encoding the UDP-MurNac-
 pentapeptide synthetase of Escherichia coli.";
 RL Nucleic Acids Res. 17:5379-5379(1989).
 RN [2]
 RN SEQUENCE FROM N.A.
 RN STRAIN-K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RT Isono K., Mizobuchi K., Nakata A.,
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [3]
 RN SEQUENCE FROM N.A.
 RN STRAIN-MB2884;
 RX MEDLINE=97128642; PubMed=8973200;
 RA Anderson M.S., Eveland S.S., Onishi H.R., Pompliano D.L.,
 RT "Kinetic mechanism of the Escherichia coli UDPMurNac-tripeptide
 D-alanyl-D-alanine-adding enzyme: use of a glutathione S-transferase
 fusion.";
 RL Biochemistry 35:16264-16269(1996).
 RN [4]
 RN SEQUENCE FROM N.A.
 RN STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blatner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RT Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [5]
 RN SEQUENCE FROM N.A., AND MUTANT MURE2.
 RN STRAIN-CGSC 5990;
 RX MEDLINE=97309380; PubMed=9166795;
 RA Eveland S.S., Pompliano D.L., Anderson M.S.;
 RT "Conditionally lethal Escherichia coli murein mutants contain point
 defects that map to regions conserved among murein and foily poly-
 gamma-glutamate ligases: identification of a ligase superfamily.";
 RL Biochemistry 36:6223-6229(1997).
 RN [6]
 RN CHARACTERIZATION, AND SEQUENCE OF 1-15.
 RX MEDLINE=90248455; PubMed=2186811;
 RA Duncan K., van Heijenoort J., Walsh C.T.;
 RT "Purification and characterization of the D-alanyl-D-alanine-adding
 enzyme from Escherichia coli.";
 RL Biochemistry 29:2379-2386(1990).
 RN [7]
 RN X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=20545602; PubMed=11090285;
 RA Yan Y., Munshi S., Leiting B., Anderson M.S., Chrzas J., Chen Z.;
 RT "Crystal structure of Escherichia coli UDPMurNac-tripeptide
 D-alanyl-D-alanine-adding enzyme (MurF) at 2.3-A resolution.";
 RL J. Mol. Biol. 304:435-445(2000).
 CC -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
 STEP IN THE SYNTHESIS OF UDP-N-ACETYL-MURAMOYL-PENTAPEPTIDE, THE
 PRECURSOR OF MUREIN.
 CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
 glutamyl-meso-2,6-diaminohexanoate + D-alanyl-D-alanine = ADP
 + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
 carboxy-L-lysyl-D-alanyl-D-alanine.
 CC -1- PATHWAY: Peptidoglycan biosynthesis.
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO THE MURCODE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: X15432; CA33473.1; -
 DR EMBL: X55034; CA38863.1; -
 DR EMBL: D10483; BAA01351.1; -
 DR EMBL: U67891; AAC4457.1; -
 DR EMBL: AE000118; AAC73197.1; -
 DR EMBL: U67893; AAB60788.1; -
 DR PIR: S04846; S04846.
 DR PIR: S40596; S40596.
 DR PDB: 1G64; 20-DEC-00.
 DR Ecocore; BG10622; murF.
 DR InterPro: IPR000713; Mur_Ligase.
 DR InterPro: IPR004101; Mur_Ligase_C.
 DR Pfam: PF02875; Mur_Ligase_C_1.
 DR TIGRFAMs: TIGR01143; murF_1.
 KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
 KW ATP-binding; 3D-structure; Complete proteome.
 FT NP_BIND 107 113 ATP (POTENTIAL).
 FT VARIANT 288 288 A -> T (IN MURE2; TS MUTANT WITH LOW
 FT ACTIVITY).
 FT CONFLICT 61 61 G -> A (IN REF. 1 AND 2).
 FT CONFLICT 178 178 A -> R (IN REF. 1 AND 2).
 SQ SEQUENCE 452 AA; 47447 MW; B46E2B57BDBBC572 CRC64;
 Query Match 25.9%; Score 58; DB 1; Length 452;
 Best Local Similarity 38.5%; Pred. No. 3.4;
 Matches 10; Conservative 7; Mismatches 5; Indels 4; Gaps 1;
 QY 12 KSGSVNAPMPNGV---TENNDDV 33
 DB 200 KAKKEIFGLPENGTAIINNADNDML 225
 ID CD72_HUMAN STANDARD; PRT; 359 AA.
 AC P21854;
 DT 01-MAY-1991 (Rel. 18, Created)
 DT 01-MAY-1991 (Rel. 18, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE B-cell differentiation antigen CD72 (Lyb-2).
 GN CD72.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=90278102; PubMed=2141045;
 RA von Hoegen I., Nakayama E., Parnes J.R.;
 RT "Identification of a human protein homologous to the mouse Lyb-2 B
 cell differentiation antigen and sequence of the corresponding
 RT cDNA.";
 RL J. Immunol. 144:4870-4877(1990).
 RN [2]
 RN SEQUENCE FROM N.A.
 RC TISSUE=Pancreas, and Spleen;
 RA Strusberg R.;
 RL Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: PLAYS A ROLE IN B CELL PROLIFERATION AND
 CC DIFFERENTIATION. ASSOCIATES WITH CDS.
 CC -1- SUBUNIT: HOMODIMER, DISULFIDE-LINKED.
 CC -1- SUBCELLULAR LOCATION: Type II membrane protein.
 CC -1- TISSUE SPECIFICITY: PRE-B-CELLS AND B-CELLS BUT NOT TERMINALLY
 CC DIFFERENTIATED PLASMA CELLS.
 CC -1- SIMILARITY: CONAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD72 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd72.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC	The European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; M54992; AAA63189.1; .-
DR	EMBL; BC030227; AAH30227.1; .-
DR	PIR; A43532; A43532.
DR	GeneW; HGNC:1696; CD72.
DR	MIM; 107272; .-
DR	InterPro; IPR001304; Lectin_C.
DR	Pfam; PF00009; lectin_c; 1.
DR	SMART; SM00034; CLECT; 1.
DR	PROSITE; PS00615; C-TYPE_LECTIN_1; FALSE_NEG.
DR	PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
DR	KW Antigen; B-cell; Glycoprotein; Transmembrane; Lectin; Signal-anchor.
FT	DOMAIN 1 95 CYTOPLASMIC (POTENTIAL)
FT	TRANSMEM 96 116 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT	(POTENTIAL).
FT	DOMAIN 117 359 EXTRACELLULAR (POTENTIAL).
FT	DOMAIN 232 352 C-TYPE LECTIN (LONG FORM).
FT	DISULFID 233 244 BY SIMILARITY.
FT	DISULFID 261 350 BY SIMILARITY.
FT	DISULFID 325 342 BY SIMILARITY.
FT	CARBOHD 136 136 N-LINKED (GLCNAC...) (POTENTIAL).
SQ	SEQUENCE 359 AA; 40220 MW; 87A52C028AC17E44 CRC64;
QY	Query Match 25.0%; Score 56; DB 1; Length 359;
Dd	Best Local Similarity 34.9%; Pred. No. 4.8;
Matches	15; Conservative 7; Mismatches 15; Indels 6; Gaps 2;
QY	1 LYXKNRYVALKS---GGSVNAPKPEGVENNDFVFMGYTQ 39
:	:::: ::: :
Dd	274 IYPQSHSYFLNSLIPNGSGNSYW--TGLSNKMDKLTDPTQ 314
RESULT 11	
LYTB_STRPN	
ID	LYTB_STRPN STANDARD; PRT; 658 AA.
AC	O9ZAP7;
DT	16-OCT-2001 (Rel. 40, Created)
DT	16-OCT-2001 (Rel. 40, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Purative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.96)
DE	(Murein hydrolase).
GN	LYTB OR SP0965.
OS	Streptococcus pneumoniae.
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC	Streptococcus.
OX	NCBI_TaxID=1313;
RN	[1]
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
RC	SPRAIN-R6.
RA	Garcia P., Gonzalez M.P., Garcia E., Lopez R., Garcia J.L.;
RT	"LybB, a novel pneumococcal murein hydrolase essential for cell
RL	separation."
RM	Mol. Microbiol. 31:1275-1281(1999).
RP	[2]
RP	SEQUENCE FROM N.A.
RC	SPRAIN-TIGR4;
RC	MEDLINE=21357209; PubMed=11463916;
RA	Tetzelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA	Petersen S., Heidelberg J., Debay R.T., Haft D.H., Dodson R.J.,
RA	Dunkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA	Umayam L.A., White O., Salzberg S.L., Lewis M.R., Redune D.,
RA	Holtzapfele E., Khouri H., Wolf A.M., Ueberback T.R., Hansen C.L.,
RA	McDonald L.A., Feldblum T.V., Angilou M.S., Dickinson T., Hickey E.K.,
RA	Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA	Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT	"Complete genome sequence of a virulent isolate of Streptococcus
RT	pneumoniae"

RU	SEQUENCE293:498-506(2001).
CC	-I- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
CC	CELL SEPARATION.
CC	-I- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC	unit in high-mannose glycoproteids and glycoproteins containing
CC	the -Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC	residue remains attached to the protein; the rest of the
CC	oligosaccharide is released intact.
CC	-I- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC	-----
CC	THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/network/or_send_an_email_to_license@isb-sib.ch).
CC	-----
DR	EMBL: AJ010312; CAA09078.1; -
DR	EMBL: AE007400; AAK75086.1; .
DR	TIGR: SP0965; .
DR	InterPro: IPR002901; Amidase_4.
DR	Pfam: PF01832; Amidase_4; 1.
DR	SMART: SM00047; LYZ2; 1.
KW	Signal; Hydrolase; Cell wall; Complete proteome.
FT	SIGNAL 1 23
FT	CHAIN 24 658
FT	PUTATIVE ENDO-BETA-N-
FT	ACETYLGALACTOSAMINIDASE.
FT	CONFLICT 336 336 I -> M (IN REF. 1).
FT	CONFLICT 381 381 T -> A (IN REF. 1).
FT	CONFLICT 384 384 E -> K (IN REF. 1).
FT	CONFLICT 535 535 L -> P (IN REF. 1).
FT	CONFLICT 580 580 F -> S (IN REF. 1).
SO	SEQUENCE 658 AA: 76469 MW: B625515006 GC3B36 CRC64;

```

Query Match          23.9%  Score 53.5;  DB 1;  Length 658;
Best Local Similarity 27.1%;  Pred. No. 20;
Matches 13;  Conservative 9;  Mismatches 15;  Indels 11;  Gaps 1;

QY      3  KNNRYVALKSGGSVNP-----MPENGVTENNDFVFMGYTQ 39
      ::  || ||||| : 1
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      213  FENGHYLLKSGGYVAANEMIMDKESWFLNFEDGKMARKWVYDSHQ 260

RESULT 12
ALYS_ENTFA
ID  Alys_ENTFA  STANDARD;  PRT;  671 AA.
P37710;
DT  01-OCT-1994 (Rel. 30, Created)
DT  01-OCT-1994 (Rel. 30, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Autolysin precursor (EC 3.2.1.-) (Peptidoglycan hydrolase)
DE  (Beta-glycosidase).
DE  Enterococcus faecalis (Streptococcus faecalis).
OC  Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
OX  NCBI_TaxID=1351;
   [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=91358349; PubMed=1679432;
RA  Bellevau C., Potvin C., Trudel J., Asselin A., Bellemare G.;
RT  "Cloning, sequencing, and expression in Escherichia coli of a
RL  Streptococcus faecalis autolysin."
RT  J. Bacteriol. 173:5619-5623(1991).
CC  -1- FUNCTION: HYDROLYZES THE CELL WALL OF E.FAECALIS AND
      M.LYSODEKTRICUS. MAY PLAY AN IMPORTANT ROLE IN CELL WALL GROWTH
      AND CELL SEPARATION.
CC  -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC  -1- DOMAIN: LYSM REPEATS ARE THOUGHT TO BE INVOLVED IN PEPTIDOLYCAN
      BINDING.
CC  -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC  -1- SIMILARITY: CONTAINS 5 LYSM REPEATS.
CC  -----
CC  This SWISS-PROT entry is copyright It is produced through a collaboration

```



```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M58002; AAA67325.1; -.
DR PIR; A38109;
DR InterPro; IPR002901; Amidase_4.
DR InterPro; IPR002482; LYSM.
DR Pfam; PF01476; LYSM; 5.
DR Pfam; PF01832; Amidase_4; 1.
DR SMART; SM00047; LY22; 1.
DR SMART; SM00257; LYSM; 5.
DR Hydroxylase; Glycosidase; Bacteriolytic enzyme; Cell wall;
KW Cell division; Septation; Repeat; Signal.
KW SIGNAL
FT CHAIN 1 53 POTENTIAL.
FT REPEAT 54 671 AUTOLYSIN.
FT REPEAT 363 405 LYSM 1.
FT REPEAT 431 473 LYSM 2.
FT REPEAT 499 541 LYSM 3.
FT REPEAT 567 609 LYSM 4.
FT REPEAT 629 671 LYSM 5.
SQ SEQUENCE 671 AA; 70442 MW; 34582686C6C1A4A33 CRC64;

Query Match 23.9%; Score 53.5; DB 1; Length 671;
Best Local Similarity 34.2%; Pred. No. 21;
Matches 13; Conservative 6; Mismatches 10; Indels 9; Gaps 1;

OY 8 YYALKSGGSVNAPENGVTENN-----DWVFMG 36
DB 430 YTVKSGDTLNTIAQGVTVANLSMNGISGDLIFVG 467

RESULT 13
RN12_YEAST STANDARD; PRT; 850 AA.
ID P32843;
AC P32843;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RN12 protein.
GN RN12 OR PRP12 OR YMR302C OR YW952.04C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28583 / FL100;
RX MEDLINE=92212295; PubMed=1557037;
RA Liang S., Alksne L., Warner J.R., Lacroute F.;
RT "RN12", a gene of Saccharomyces cerevisiae involved in pre-RNA
RT maturation. Characterization of a temperature-sensitive mutant,
RT cloning and sequencing of the gene."
RL Mol. Genet. 232:304-312(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RT Submitted (Apr-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PLAYS A DISPENSABLE ROLE IN EARLY MATURATION OF
CC PRE-RRNA. IT MAY PROVIDE A COFACTOR WHICH PLAYS A MINOR AND
CC DISPENSABLE ROLE IN PRE-RRNA MATURATION. THIS PROTEIN MIGHT
CC MEDIATE PROCESSIVELY ALONG EACH PRE-RNA MOLECULE IN THE EARLY
CC STEPS OF THE RNA MATURATION PATHWAY AND BE RELEASED ONCE THE
CC MODIFICATION IT CATALYSES HAD BEEN COMPLETED.
CC -1- SIMILARITY: TO S.POMBE SPBC83.05.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; S92205; AAB2191.1; -.
DR EMBL; 249212; CAA89135.1; -.
DR PIR; S20462; S20462.
DR SGD; S0004917; PRP12.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF00076; rrm; 1.
DR SMART; SM00360; RRM; 1.
DR VARIANT 502 502
FT
FT
FT
FT
SQ SEQUENCE 850 AA; 96688 MW; 3BEFD30AAD376C7 CRC64;

Query Match 23.9%; Score 53.5; DB 1; Length 850;
Best Local Similarity 44.8%; Pred. No. 27;
Matches 13; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

OY 4 KNRYTALKS-GGSVNAPENGVTENNND 31
DB 597 EDLYYNNKSKGENVKEPESEKETAEENNND 625

RESULT 14
TRDN_CANFA STANDARD; PRT; 700 AA.
ID TRDN_CANFA
AC P82179;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Triadin.
GN TRDN.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC TISSUE=Heart, and skeletal muscle;
RX MEDLINE=99428545; PubMed=10497235;
RA Kobayashi Y.M., Jones L.R.;
RT "Identification of triadin 1 as the predominant triadin isoform
RT expressed in mammalian myocardium";
RL J. Biol. Chem. 274:28660-28668(1999).
CC -1- FUNCTION: MAY BE INVOLVED IN ANCHORING CALDESQUESTRIN TO THE
CC JUNCTIONAL SARCOPLASMIC RETICULUM AND ALLOWING ITS FUNCTIONAL
CC COUPLING WITH THE RYANODINE RECEPTOR (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; CARDIAC ISOFORM 1, SKELETAL
CC ISOFORM (SHOWN HERE) AND CARDIAC ISOFORM 3; ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: SKELETAL AND CARDIAC MUSCLE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF165916; AAF00222.1; -.
DR EMBL; AF165915; AAF00221.1; -.
DR EMBL; AF165917; AAF00223.1; -.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW Alternative splicing.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 46 CYTOPLASMIC (POTENTIAL).

```



```

ET      CONFLICT      56      56      G -> A (IN REF. 21).
FT      CONFLICT      251
SQ      SEQUENCE      417 AA; 45027 MW; CDBCF0054AE2FDO CRC64;

Query Match      23.2%; Score 52; DB 1; Length 417;
Best Local Similarity 36.7%; Pred. No. 19;
Matches 11; Conservative 6; Mismatches 11; Indels 2; Gaps 1;

OY      12 KSGGSVNAPMPEN--GVTENDDWEMGYTQ 39
      : | : | : | | | : | |
Db      183 QOGAIIAAAIPTSRAGVTPNDMDAIIQMRQ 212

Search completed: December 10, 2002, 10:54:39
Job time : 7.6 secs

```

Search completed: December 10, 2002, 10:54:39
Job time : 7.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:49:03 ; Search time 21.8 Seconds
(without alignments)
378.068 Million cell updates/sec

Title: US-09-142-970-2_COPY_1_40
Perfect score: 224
Sequence: 1 LYKKNRYALKSGSVNAPMPENGVTENDWVEMGYTQE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	224	100.0	993	2	Q9S6X3
2	224	100.0	996	2	Q30574
3	224	100.0	997	2	Q9S6X2
4	224	100.0	1552	2	Q9A106
5	212	94.6	996	2	Q57309
6	202	90.2	992	2	Q57035
7	202	90.2	992	2	Q9S6X5
8	202	90.2	992	2	Q9S6X4
9	202	90.2	992	2	Q30575
10	202	90.2	997	2	Q30575
11	202	90.2	1561	2	Q51169
12	202	90.2	1773	16	Q9JVB9
13	198	88.4	1815	16	Q9K0B4
14	100	44.6	1764	2	Q9J3T4
15	59.5	26.6	130	2	Q9ZAF6
16	58	25.9	452	16	Q8R06

17	58	25.9	452	16	Q829H2	Q829H2 salmonella
18	58	25.9	452	16	Q8X9Z1	Q8X9Z1 escherichia
19	56	25.0	473	16	Q8RDP2	Q8RDP2 fusobacteri
20	56	25.0	552	17	Q58688	Q58688 metanococc
21	56	25.0	574	16	Q8XEC2	Q8XEC2 escherichia
22	55	25.0	1367	5	Q9RYU0	Q9RYU0 caenorhabdi
23	55	24.6	151	12	P89289	P89289 xestia c-ni
24	55	24.6	217	17	Q96XX6	Q96XX6 sulfobolus
25	55	24.6	283	9	Q9AF60	Q9AF60 streptococc
26	55	24.6	739	5	Q9N8U6	Q9N8U6 trypanosoma
27	55	24.6	859	12	Q9PYP8	Q9PYP8 xestia c-ni
28	55	24.6	2316	2	Q9PFD9	Q9PFD9 bacteroides
29	54.5	24.3	175	5	Q8SM64	Q8SM64 encephalito
30	54	24.1	350	2	Q43983	Q43983 acinetobact
31	54	24.1	353	2	Q32433	Q32433 acinetobact
32	54	24.1	577	16	Q92FP4	Q92FP4 listeria in
33	53.5	23.9	455	5	Q24464	Q24464 dirosophila
34	53.5	23.9	459	5	Q9NGW6	Q9NGW6 dirosophila
35	53.5	23.9	459	5	Q9VW46	Q9VW46 dirosophila
36	53.5	23.9	463	5	Q9NBK5	Q9NBK5 dirosophila
37	53.5	23.9	601	16	Q98BM1	Q98BM1 rhizobium 1
38	53.5	23.9	614	2	Q9AHT8	Q9AHT8 streptococc
39	53.5	23.9	648	2	Q48040	Q48040 haemophilus
40	53.5	23.9	766	5	Q18673	Q18673 caenorhabdi
41	53	23.7	454	5	Q8WPD9	Q8WPD9 branchiosto
42	53	23.7	801	16	Q8XV55	Q8XV55 ristonla s
43	53	23.7	2146	13	Q9PUM3	Q9PUM3 gallus gall
44	53	23.7	2157	13	Q9PW44	Q9PW44 gallus gall
45	53	23.7	2171	13	Q9PW46	Q9PW46 gallus gall

ALIGNMENTS

RESULT 1	ID	Q9S6X3	PRELIMINARY:	PRT:	993 AA.
AC	Q9S6X3				
DT	01-MAY-2000	(TREMBLrel. 13, Created)			
DT	01-MAY-2000	(TREMBLrel. 13, Last sequence update)			
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)			
DE	IGal protease (Fragment).				
GN	IGA.				
OS	Neisseria meningitidis.				
OC	Bacteria, Proteobacteria: beta subdivision: Neisseriaceae: Neisseria.				
OX	NCBI_TaxID=487;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=24400;				
RX	MEDLINE=98010345; PubMed=9350862;				
RA	Morelli G., Malorny B., Muller K., Seltzer A., Wang J.F., del Valle J.,				
RT	Achtmann M.;				
RT	"Clonal descent and microevolution of Neisseria meningitidis during 30				
RT	years of epidemic spread."				
RL	Mol. Microbiol. 25:1047-1064(1997).				
DR	EMBL: AF012208; AAC45791.2; --				
DR	MEROPS: S06.001; --				
DR	InterPro: IPR002195; Dihydrocoralase.				
DR	InterPro: IPR000710; IGA_S6.				
DR	InterPro: IPR004899; Pertactin_sup.				
DR	Pfam: PF02395; IGA1.1.				
DR	Pfam: PF03212; Pertactin.1.				
DR	PRINTS: PR00921; IGASERTPASE.				
DR	PROSITE: PS00482; DIHYDROCORALASE_1; UNKNOWN_1.				
KW	PROtease.				
FT	NON_TER				
FT	NON_TER				
FT	SEQUENCE	993 AA; 109441 MM; 109FAA2EF88AC3C6 CRC64;			
SO	SEQUENCE	993 AA; 109441 MM; 109FAA2EF88AC3C6 CRC64;			
Query Match		100.0%; Score 224; DB 2; Length 993;			
Best Local Similarity		100.0%; Pred. No. 2.2e-20;			
Matches	40; Conservative	0; Mismatches	0; Indels	0; Gaps	0;

OY 1 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 40
|||||
DB 558 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 597

RESULT 2

ID 030574 PRELIMINARY; PRT; 996 AA.
AC 030574;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGAI protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24024;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread."
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012206; AAC45789.2; -
DR MEROPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGAI; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1 1
FT NON_TER 996 996
SQ SEQUENCE 996 AA; 109688 MW; B3ABDEF54C33709 CRC64;

Query Match 100.0%; Score 224; DB 2; Length 996;
Best Local Similarity 100.0%; Pred. No. 2.2e-20;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 40
|||||
DB 557 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 596

RESULT 3
OY 0956X2
ID 0956X2 PRELIMINARY; PRT; 997 AA.
AC 0956X2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGAI protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24081;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread."
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012209; AAC45792.2; -
DR MEROPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.

DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGAI; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1 1
FT NON_TER 997 997
SQ SEQUENCE 997 AA; 109812 MW; 06F2E361E7E202E0 CRC64;

Query Match 100.0%; Score 224; DB 2; Length 997;
Best Local Similarity 100.0%; Pred. No. 2.2e-20;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 40
|||||
DB 558 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 597

RESULT 4

ID 09AIU6 PRELIMINARY; PRT; 1552 AA.
AC 09AIU6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGAI protease.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RA Vitovskis, Sayers J.R.;
RT "Degenerate specificity of Neisseria meningitidis IGAI protease."
RT Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RL EMBL: AF235032; AAK150B3.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGAI; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
SQ SEQUENCE 1552 AA; 171205 MW; E49D6657B120CEFF CRC64;

Query Match 100.0%; Score 224; DB 2; Length 1552;
Best Local Similarity 100.0%; Pred. No. 3.7e-20;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 40
|||||
DB 574 LYKKNRYALKSGGSVNAMPENGVTENNDDVFMGYTOE 613

RESULT 5

ID 057309 PRELIMINARY; PRT; 996 AA.
AC 057309;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGAI protease precursor (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23524;

RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achman M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread";
 RL Mol. Microbiol. 25:1047-1064(1997).
 (12)
 RP SEQUENCE OF 37-532 FROM N.A.
 RC STRAIN-ETH2;
 RX MEDLINE-95302961; PubMed-7783620;
 RA Lomholt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding Iga1 protease
 RT in *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Haemophilus*
 RT influenzae";
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL: AF012207; AAC45790.2; -;
 DR EMBL: X82469; CAA57852.1; -;
 DR EMBL: X82468; CAA57851.1; -;
 DR MEROPS: S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1
 KM Protease.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 996 AA; 109717 MW; AB39F7B98A41F986 CRC64;

Query Match 94.6%; Score 212; DB 2; Length 996;
 Best Local Similarity 95.0%; Pred. No. 7.9e-19;
 Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 40
 DB 557 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 596

RESULT 6
 ID Q57035 PRELIMINARY; PRT; 992 AA.
 AC Q57035;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Iga1 protease precursor (Fragment).
 GN IGA.
 OS *Neisseria meningitidis*.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_TaxId=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B40;
 RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F.,
 RA del Valle J., Achman M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread";
 RL Mol. Microbiol. 25:1047-1064(1997).
 (12)
 RP SEQUENCE OF 37-532 FROM N.A.
 RC STRAIN-HP48;
 RX MEDLINE-95302961; PubMed-7783620;
 RA Lomholt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding Iga1 protease
 RT in *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Haemophilus*
 RT influenzae";
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL: AF012211; AAC45794.2; -;
 DR EMBL: X82480; CAA57863.1; -;
 DR EMBL: X82475; CAA57858.1; -;

DR MEROPS: S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KM Protease.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 992 AA; 109228 MW; 3677DDEACE6D9F69 CRC64;

Query Match 90.2%; Score 202; DB 2; Length 992;
 Best Local Similarity 90.0%; Pred. No. 1.6e-17;
 Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 40
 DB 557 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 596

RESULT 7
 ID Q9S6X5 PRELIMINARY; PRT; 992 AA.
 AC Q9S6X5;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Iga1 protease (Fragment).
 GN IGA.
 OS *Neisseria meningitidis*.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_TaxId=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-42491;
 RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achman M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread";
 RL Mol. Microbiol. 25:1047-1064(1997).
 DR EMBL: AF012203; AAC45786.1; -;
 DR MEROPS: S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KM Protease.
 FT NON_TER
 FT NON_TER
 SQ SEQUENCE 992 AA; 109239 MW; 306FF451B7213DA9 CRC64;

Query Match 90.2%; Score 202; DB 2; Length 992;
 Best Local Similarity 90.0%; Pred. No. 1.6e-17;
 Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 40
 DB 557 LYKNNRYALKSGSVNAPMPENGVTENNDDVFMGYTQE 596

RESULT 8
 ID Q9S6X4 PRELIMINARY; PRT; 992 AA.
 AC Q9S6X4;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J., Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012204; AAC45787.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CE3 CRC64;

Query Match 90.2%; Score 202; DB 2; Length 992;
Best Local Similarity 90.0%; Pred. No. 1.6e-17;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNNDWVFNGYTOE 40
DB 557 LYYKNRYVALKSGSVNAPMPENGVTENNNDWILMGSTQE 596
|||||
ID 030573 PRELIMINARY; PRT; 992 AA.
AC 030573;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23910;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J., Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012205; AAC45788.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109255 MW; EF7CA563BC7D7ECB CRC64;

Query Match 90.2%; Score 202; DB 2; Length 992;

Best Local Similarity 90.0%; Pred. No. 1.6e-17;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNNDWVFNGYTOE 40
DB 557 LYYKNRYVALKSGSVNAPMPENGVTENNNDWILMGSTQE 596
|||||
ID 030575 PRELIMINARY; PRT; 997 AA.
AC 030575;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24099;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J., Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012210; AAC45793.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF03212; Pertactin; 1.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 997
SQ SEQUENCE 997 AA; 110152 MW; 4DE40DE594E5E28E CRC64;

Query Match 90.2%; Score 202; DB 2; Length 997;
Best Local Similarity 90.0%; Pred. No. 1.6e-17;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGSVNAPMPENGVTENNNDWVFNGYTOE 40
DB 558 LYYKNRYVALKSGSVNAPMPENGVTENNNDWILMGSTQE 597
|||||
ID 051169 PRELIMINARY; PRT; 1561 AA.
AC 051169;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE IGA1 protease.
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HP13;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IGA1 protease in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus influenzae."
RL Mol. Microbiol. 15:495-506(1995).

DR EMBL: X82474; CAA57857.1; -.
 DR MEROPS: S06.001; -.
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 2.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Protease.
 SO SEQUENCE 1561 AA; 171849 MW; 1C96E291A00017D5 CRC64;

Query Match 90.2%; Score 202; DB 2; Length 1561;
 Best Local Similarity 90.0%; Pred. No. 2,6e-17;
 Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGSVNAPMPENGVTENNNDWFMGYTOE 40
 DB 584 LYKKNRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 623

RESULT 12

O9YVB9 PRELIMINARY: PRT; 1773 AA.
 AC O9YVB9;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE IGA1 protease (EC 3.4.21.7).
 GN IGA OR NMA0905.
 OS *Neisseria meningitidis* (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=65699;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222356; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S.,
 Jagsels K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skellon J.,
 Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491."
 RL Nature 404:502-506(2000).
 DR EMBL: AL162754; CAB84182.1; -.
 DR MEROPS: S06.001; -.
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 2.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Protease; Hydrolase; Complete proteome.
 SO SEQUENCE 1773 AA; 196350 MW; CAC19E713D76CE1 CRC64;

Query Match 90.2%; Score 202; DB 16; Length 1773;
 Best Local Similarity 90.0%; Pred. No. 3,1e-17;
 Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGSVNAPMPENGVTENNNDWFMGYTOE 40
 DB 573 LYKKNRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 612

RESULT 13

O9KOB4 PRELIMINARY: PRT; 1815 AA.
 AC O9KOB4;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE IGA-specific serine endopeptidase.
 GN NMB0700.
 OS *Neisseria meningitidis* (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=491;
 OX [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettein H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 Eissen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
 Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 Mason T., Cleecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,
 Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 Gill J., Scariato V., Maignan V., Pizzi M., Grandi G., Sun L.,
 Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
 RT "Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58."
 RL Science 287:1809-1815(2000).
 DR EMBL: AE002424; AAF41117.1; -.
 DR MEROPS: S06.001; -.
 DR TIGR: NMB0700;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Complete proteome.
 SO SEQUENCE 1815 AA; 201077 MW; 2259D4D7176C57F CRC64;

Query Match 88.4%; Score 198; DB 16; Length 1815;
 Best Local Similarity 87.5%; Pred. No. 1e-16;
 Matches 35; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 LYKKNRYALKSGSVNAPMPENGVTENNNDWFMGYTOE 40
 DB 573 LYKKNRYALKSGSVNAPMPENGOTENNNDWILMGSTOE 612

RESULT 14

O93T34 PRELIMINARY: PRT; 1764 AA.
 AC O93T34;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE IGA1 protease type 2.
 GN IGA1.
 OS *Haemophilus aegyptius*.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 NCBI_TaxID=725;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=F3031;
 RA McGillivray G., Actis L.A.;
 RT "Iga protease from *H. aegyptius* F3031."
 RL Submitted (Apr-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AF369907; AAK56925.1; -.
 DR MEROPS: S06.001; -.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertact-sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR Protease.
 SO SEQUENCE 1764 AA; 193863 MW; EC583CDBE81DBEC CRC64;

Query Match 44.6%; Score 100; DB 2; Length 1764;
 Best Local Similarity 44.7%; Pred. No. 0.00052;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:14:59 ; Search time 28.2 Seconds
(without alignments)
189.008 Million cell updates/sec

Title: US-09-142-970-3_COPY_1_40
Perfect score: 225
Sequence: 1 LTKKNRYALKSGSVNAPMPENGQTEINDVFMGKOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002:*

1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*

2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*

3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*

4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*

5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*

6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*

7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*

8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*

9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*

10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*

11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*

12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*

13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*

14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*

15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*

16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*

17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*

18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*

19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*

20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*

21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*

22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	225	100.0	104	19	AAW61604
2	212	94.2	104	19	AAW61603
3	203	90.2	105	19	AAW63656
4	203	90.2	105	19	AAW61602
5	199	88.4	104	19	AAW61605
6	198	88.0	104	11	AAW61606
7	94	41.8	1541	11	AAW07304
8	57.5	25.6	1364	22	AAW70912
9	55.5	24.7	569	19	AAW61214
10	55.5	24.7	569	23	ABP54633

11	55.5	24.7	591	19	AAW55099
12	55.5	24.7	591	23	ABP54593
13	55.5	24.7	599	21	AAW81517
14	55.5	24.7	678	21	AAW81667
15	55	24.4	452	22	AAU34434
16	55	24.4	452	22	AAU38481
17	55	24.4	452	22	AAW698402
18	54.5	24.2	948	22	ABP58055
19	53.5	23.8	1317	22	AAU33622
20	53	23.6	479	22	ABG25468
21	53	23.6	845	22	ABG28778
22	53	23.6	881	22	ABG24987
23	52.5	23.3	194	23	ABP07412
24	52.5	23.3	456	17	AAW83655
25	52.5	23.3	459	22	ABW71664
26	52	23.1	392	22	AAE07963
27	52	23.1	392	22	AAU04473
28	52	23.1	392	22	AAU04614
29	52	23.1	392	22	AAU01868
30	52	23.1	392	22	AAU00267
31	51.5	22.9	317	23	AAU93165
32	51.5	22.9	671	17	AAW85290
33	51.5	22.9	862	22	AAW78425
34	51.5	22.9	863	22	AAW03196
35	51.5	22.9	895	22	AAW79409
36	51	22.7	604	20	AAW93168
37	51	22.7	641	15	AAW62936
38	50.5	22.4	340	21	AAW33319
39	50.5	22.4	454	22	AAW78021
40	50.5	22.4	465	21	AAW19011
41	50.5	22.4	465	22	AAW78018
42	50.5	22.4	465	22	AAW19996
43	50.5	22.4	465	22	AAW19998
44	50.5	22.4	479	21	AAW19284
45	50.5	22.4	479	21	AAW43132

ALIGNMENTS

RESULT 1

AAW61604

ID AAW61604 standard; peptide: 104 AA.

XX

AC AAW61604:

DT 27-OCT-1998 (first entry)

XX

DE Neisseria IgA1 protease fragment 3.

XX

KW Immunoglobulin protease; carrier; paediatric; vaccine;

KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.

OS Neisseria sp.

XX

PN W09831791-A1.

XX

PD 23-JUL-1998.

XX

PF 20-JAN-1998; 98WO-EP00294.

XX

PR 21-JAN-1997; 97EP-0100883.

XX

PA (PLAC) MAX PLANCK GES. FORERDERUNG WISSENSCHAFTEN.

PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.

PI Achtmann M, Moreau M;

XX

DR WPI; 1998-414092/35.

XX

PT New peptide from Neisseria immunoglobulin protease - useful as

PT immunogenic carrier, e.g. particularly for polysaccharide(s),

PT forming conjugates used in vaccines against Neisseria and

PT Haemophilus
XX
PS Claim 2: Fig 1: 32pp: English.

CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.

XX Sequence 104 AA:

QY Best Local Similarity 100.0%; Score 225; DB 19; Length 104;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1 LYYKNRYALKSGSVNAPMENGQTEENNDDWFMGYKOE 40
1 LYYKNRYALKSGSVNAPMENGQTEENNDDWFMGYKOE 40

RESULT 2

AAW61603
ID AAW61603 standard; peptide; 104 AA.

XX AAW61603;

DT 27-OCT-1998 (first entry)

DE Neisseria IgA1 protease fragment 2.

XX Immunoglobulin protease; carrier; paediatric; vaccine;
KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.

OS Neisseria sp.

PN WO9831791-A1.

PD 23-JUL-1998.

PF 20-JAN-1998; 98WO-EP00294.

PR 21-JAN-1997; 97EP-0100883.

PA (PLAC) MAX PLANCK GES. FORDERUNG WISSENSCHAFTEN.
PI (INNR) PASTEUR MERIEUX SERUMS & VACCINS SA.

PI Achtmann M, Moreau M;

DR WPI: 1998-414092/35.

PT New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus

PS Claim 2: Fig 2: 32pp: English.

XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.

XX Sequence 104 AA:

Query Match 94.2%; Score 212; DB 19; Length 104;

Best Local Similarity 95.0%; Pred. No. 1.3e-22;
Matches 38; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LYYKNRYALKSGSVNAPMENGQTEENNDDWFMGYKOE 40
1 LYYKNRYALKSGSVNAPMENGQTEENNDDWFMGYKOE 40

RESULT 3

AAW65656
ID AAW65656 standard; peptide; 105 AA.

XX AAW65656;

DT 15-OCT-1998 (first entry)

DE 105-mer peptide used in polysaccharide-peptide conjugate.

XX Polysaccharide-peptide conjugate; vaccine; linker moiety; adjuvant;

XX Immune response.

OS Synthetic.

XX Key Location/Qualifiers

FT Modified-site 1 /note="N-terminal acetyl"

PN WO9831393-A2.

PD 23-JUL-1998.

PF 21-JAN-1998; 98WO-EP00654.

PR 21-JAN-1997; 97EP-0100884.

PA (INNR) PASTEUR MERIEUX SERUMS & VACCINS SA.

PI Mistretta N, Moreau M;

DR WPI: 1998-413820/35.

PT Polysaccharide-peptide conjugates, useful as, e.g. vaccines -
PT comprise peptide moiety with at least six amino acid residues,
PT polysaccharide chain with at least four repeat units, and linker
PT moiety

PS Example 1; Page 14; 28pp: English.

XX The invention relates to: (A) polysaccharide-peptide conjugate (in which
CC the polysaccharide is immunogenic), comprising: (a) a peptide moiety
CC which has at least 6 amino acid residues, at least 1 of which is a
CC cysteine residue; (b) a polysaccharide chain comprising at least 4 repeat
CC units, and (c) a linker moiety bound to the thiol group of the cysteine.
CC The linker is also bound to: (i) native amino, hydroxyl or carboxyl
CC groups of the polysaccharide chain; (ii) amino groups created by
CC hydrolysis of native N-acyl groups of the polysaccharide chain, or (iii)
CC functional groups introduced on the polysaccharide chain upon
CC derivatisation with a spacer moiety bound to native amino, hydroxyl or
CC carboxyl groups of the polysaccharide chain, and (B) conjugating a
CC peptide as in (Aa) to a polysaccharide chain comprising at least 4 repeat
CC units, comprising: (a) coupling the peptide to a linker through the thiol
CC group of the cysteine residue, and (b) coupling the linker to the
CC polysaccharide chain through one of groups (i), (ii) or (iii) as
CC described in (A). The conjugates are especially useful as vaccines to
CC elicit a protective long term immune response against a pathogenic
CC microorganism from which the immunogenic polysaccharide chain is derived.
CC Known polysaccharide-peptide conjugates are less immunogenic than their
CC polysaccharide-protein counterparts and require adjuvantation. The new
CC conjugates have good immunogenicity, largely because the peptides are of
CC sufficient size. The present sequence represents a peptide used in
CC a polysaccharide-peptide conjugate.

XX Sequence 105 AA:


```

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (IMMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
PI Achtmann M, Moreau M;
XX
DR WPI: 1998-414092/35.
XX
PT New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
PS Claim 2; Fig 4; 32pp; English.
XX
CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
CC
XX Sequence 104 AA;
SQ
Query Match 88.0%; Score 198; DB 19; Length 104;
Best Local Similarity 85.0%; Pred. No. 1.4e-20;
Matches 34; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
OY 1 LYKNRYRYALKSGSVNAPMPENGQTEENNDWVPMYKQOE 40
    |||||
DB 1 LYKNRYRYALKSGGRLNAPMPENGVAENNDWIFMGYQOE 40
    |||||

RESULT 7
AAW07304
ID AAR07304 standard; protein: 1541 AA.
XX
AC AAR07304;
XX
DT 31-JAN-1991 (first entry)
XX
DE Iga1 protease.
XX
KM Iga1; vaccine; meningitis; gonorrhoea; allergies.
XX
OS Haemophilus influenzae.
XX
PN WO9011367-A.
XX
PD 04-OCT-1990.
XX
PF 16-MAR-1990; 90WO-DK00073.
XX
PR 17-MAR-1989; 89DK-0001308.
XX
PA (KIL1/) KILIAN M.
XX
PI Kilian M, Poulsen K;
XX
DR WPI: 1990-320267/42.
XX
DR N-PSDB; AAQ06164.
XX
PT Immunoglobulin A1 protease prodn. - by cloning from
PT microorganisms for immunisation against immunoglobulin A1
PT protease producing bacteria
XX
PS Disclosure; fig 3; 44pp; English.
XX
CC This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,
CC gonorrhoea or allergic diseases. It specifically cleaves the heavy

```

```

CC chain of human Iga1 in the hinge region.
XX
SQ Sequence 1541 AA;
XX
Query Match 41.8%; Score 94; DB 11; Length 1541;
Best Local Similarity 41.7%; Pred. No. 0.00029;
Matches 15; Conservative 8; Mismatches 13; Indels 0; Gaps 0;
OY 1 LYKNRYRYALKSGSVNAPMPENGQTEENNDWVPMG 36
    |::| |||||::|::|:|:|:|:|:|:|:|:|
DB 595 LNLNTYIALRKGASTRSELPRNKGSESNWLYMG 630
    |::| |||||::|::|:|:|:|:|:|:|:|:|

RESULT 8
AAW070912
ID AAG70912 standard; protein: 1364 AA.
XX
AC AAG70912;
XX
DT 27-JUL-2001 (first entry)
XX
DE S cerevisiae apoptosis associated protein YER183C.
XX
KM Yeast; fungus; apoptosis; infection; proliferative disease;
KM vaccine; autoimmune disease; ischaemia; neurodegeneration.
XX
OS Saccharomyces cerevisiae.
XX
PN WO200102550-A2.
XX
PD 11-JAN-2001.
XX
PF 03-JUL-2000; 2000WO-BE00077.
XX
PR 01-JUL-1999; 99EP-0870141.
XX
XX (JANC ) JANSSEN PHARM NV.
XX
PI Contreras RH, De Backer MD, Luyten WHML, Malcorps IKL;
XX
PI Nellissen BJM, Reekmans RJ;
XX
DR WPI: 2001-367042/38.
XX
DR N-PSDB; AAH29948.
XX
PT Yeast and fungal nucleic acids encoding proteins involved in a pathway
PT leading to programmed cell death, useful for treating proliferative
PT disorders, yeast and fungal infections, or for preventing apoptosis in
PT certain diseases -
XX
PS Claim 1; Fig 1; 218pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of apoptosis associated proteins from the yeast Saccharomyces
CC cerevisiae and the fungus Candida albicans. These can be used to identify
CC treatments for fungal and yeast infections, for proliferative diseases
CC and for apoptosis related diseases such as autoimmune diseases, ischaemia
CC and neurodegeneration. The present sequence is one of the S. cerevisiae
CC proteins of the invention.
XX
SQ Sequence 1364 AA;
XX
Query Match 25.6%; Score 57.5; DB 22; Length 1364;
Best Local Similarity 37.9%; Pred. No. 42;
Matches 11; Conservative 6; Mismatches 7; Indels 5; Gaps 1;
OY 3 YKNRYRYALKSGSVNAPMPENGQTEENND 31
    |::|:|:| ||||| | | | |
DB 519 YRSHRFISIEKGSVG-----NSNTGND 542
    |::|:|:| ||||| | | | |

RESULT 9
AAW61214
ID AAW61214 standard; protein: 569 AA.

```

```
XX AC AAW61214;
XX XX
XX DT 02-OCT-1998 (first entry)
XX DE Streptococcus pneumoniae SP0089 protein.
XX KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX XX detection; pneumonia; otitis media; meningitis.
XX OS Streptococcus pneumoniae.
XX PN WO9818930-A2.
XX PD 07-MAY-1998.
XX PF 30-OCT-1997; 97WO-US19422.
XX PR 31-OCT-1996; 96US-0029960.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Choi GH, Hiromocky J A, Johnson LS, Kunsch CA;
XX DR WPI: 1998-272224/24.
XX DR N-PSDB; AAV27400.
XX PT Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX PS pneumoniae - or their epitope-containing fragments, useful in
XX XX protective or therapeutic vaccines, and for diagnosis
XX XX
XX PS Claim 11; Page 80; 118pp; English.
XX CC The present sequence represents a protein from Streptococcus pneumoniae.
XX CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX CC can be useful in vaccines for inducing protective antibodies against
XX CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX CC are used to detect Streptococcus infection (by usual hybridisation or
XX CC amplification methods), also for isolating Streptococcus genes or their
XX CC allelic variants. The protein can be used similarly to detect specific
XX CC antibodies in standard immunoassays, especially for diagnosing or
XX CC monitoring infections. Antibodies which bind the protein are used to
XX CC detect corresponding antigens, to purify the protein and for passive
XX CC immunisation (optionally coupled to a toxin). Vaccines are administered,
XX CC e.g. by injection, orally or through the skin, typically at 0.01-1000
XX CC (especially 10-300) mu g/ml per dose.
XX SQ Sequence 569 AA;
XX
XX Query Match 24.7%; Score 55.5; DB 19; Length 569;
XX Best Local Similarity 27.1%; Pred. No. 29;
XX Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;
OY 3 YKNRYVALKSGGVNAP-----MPENGQTEINDWVEMGYKQ 39
XX : : | | | | | : | : : : | : |
DB 129 FENGHYIYLLKSGGYMANEMWMDKESWFLKFDGKMAEKEMVYDSHSQ 176
XX
XX RESULT 10
XX ABP54633
XX ID ABP54633 standard; Protein; 569 AA.
XX AC ABP54633;
XX XX
XX DT 04-SEP-2002 (first entry)
XX XX
XX DE S. pneumoniae SP089 protein sequence SEQ ID NO:154.
XX XX
XX KM Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
XX XX antibacterial; Streptococcal infection; detection.
XX OS Streptococcus pneumoniae.
```

```
XX XX US2002061545-A1.
XX PN
XX XX 23-MAY-2002.
XX PD
XX XX 22-JAN-2001; 2001US-0765272.
XX PF
XX PR 30-OCT-1997; 97US-0961083.
XX XX
XX PA (CHOI/) CHOI G H.
XX PA (KUNS/) KUNSCH C A.
XX PA (BARA/) BARASH S C.
XX PA (DILL/) DILLON P J.
XX PA (DOUG/) DOUGHERTY B.
XX PA (FANN/) FANNON M R.
XX PA (ROSE/) ROSEN C A.
XX XX
XX PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX PI Rosen CA;
XX DR WPI: 2002-479261/51.
XX DR N-PSDB; AB084868.
XX XX
XX PT New Streptococcus pneumoniae antigens, useful for detecting
XX PT Streptococcus and for preventing or attenuating disease caused by
XX PT Streptococcus infection -
XX XX
XX PS Claim 11; Page 41-42; 70pp; English.
XX CC AB08492 to AB084904 represents nucleic acids which encode the
XX CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
XX CC The S. pneumoniae antigens have antibacterial activity and can be
XX CC used in vaccines. The S. pneumoniae antigens can also be used to
XX CC prevent or attenuate a Streptococcal infection in an animal. The
XX CC polynucleotides encoding the S. pneumoniae antigens can be used to
XX CC detect Streptococcus nucleic acids. AB084905 to AB085130 represent
XX CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX CC which are used in an example from the present invention.
XX SQ Sequence 569 AA;
XX
XX Query Match 24.7%; Score 55.5; DB 23; Length 569;
XX Best Local Similarity 27.1%; Pred. No. 29;
XX Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;
OY 3 YKNRYVALKSGGVNAP-----MPENGQTEINDWVEMGYKQ 39
XX : : | | | | | : | : : : | : |
DB 129 FENGHYIYLLKSGGYMANEMWMDKESWFLKFDGKMAEKEMVYDSHSQ 176
XX
XX RESULT 11
XX AAW55099
XX ID AAW55099 standard; Protein; 591 AA.
XX AC AAW55099;
XX XX
XX DT 02-OCT-1998 (first entry)
XX XX
XX DE Streptococcus pneumoniae SP0046 protein.
XX XX
XX KM Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX KM detection; pneumonia; otitis media; meningitis.
XX OS Streptococcus pneumoniae.
XX FH
XX XX Key Location/Qualifiers
XX FT Misc-difference 24 /label= unknown
XX FT /note="encoded by TTN"
XX PN WO9818930-A2.
XX PD 07-MAY-1998.
```

XX 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromocky J A, Johnson LS, Kunsch CA;
XX WPI: 1998-272224/24.
XX N-PSDB; AAV27360.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX pneumoniae - or their epitope-containing fragments, useful in
XX protective or therapeutic vaccines, and for diagnosis
XX
XX Claim 11; Page 64; 118pp; English.
XX
XX The present sequence represents a protein from Streptococcus pneumoniae.
XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX can be useful in vaccines for inducing protective antibodies against
XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX are used to detect Streptococcus infection (by usual hybridisation or
XX amplification methods), also for isolating Streptococcus genes or their
XX allelic variants. The protein can be used similarly to detect specific
XX antibodies in standard immunoassays, especially for diagnosing or
XX monitoring infections. Antibodies which bind the protein and for passive
XX immunisation (optionally coupled to a toxin). Vaccines are administered,
XX e.g. by injection, orally or through the skin, typically at 0.01-1000
XX (especially 10-300) mu g/ml per dose.
XX
XX Sequence 591 AA:

Query Match 24.7%; Score 55.5; DB 19; Length 591;
Best Local Similarity 27.1%; Pred. No. 30;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;
QY 3 YKNRYRYALKSGSVNAP-----MPENGQTEINDWVFMGYKQ 39
DB 190 FENGHHYYLKSGGYMANEMIMDKESWFLKFDGKMAEKEMWYDSSHQ 237

RESULT 12
ABP54593
ID ABP54593 standard; Protein; 591 AA.
XX
XX ABP54593;
XX
XX 04-SEP-2002 (first entry)
XX
XX S. pneumoniae SP046 protein sequence SEQ ID NO:74.
XX
XX Streptococcus pneumoniae: epitope; vaccine; antigenic protein;
XX antibacterial; Streptococcal infection; detection.
XX
XX Streptococcus pneumoniae.
XX
XX US2002061545-A1.
XX
XX 23-MAY-2002.
XX
XX 22-JAN-2001; 2001US-0765272.
XX
XX 30-OCT-1997; 97US-0961083.
XX
XX (CHOI/) CHOI G H.
XX (KUNS/) KUNSCH C A.
XX (BARA/) BARASH S C.
XX (DILL/) DILLON P J.
XX (DOUG/) DOUGHERTY B.
XX (FANN/) FANNON M R.
XX

PA (ROSE/) ROSEN C A.
XX
XX Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
XX Rosen CA;
XX
XX WPI: 2002-479261/51.
XX N-PSDB; AB084828.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
XX Streptococcus and for preventing or attenuating disease caused by
XX Streptococcus infection -
XX
XX Claim 11; Page 30-31; 70pp; English.
XX
XX AB084792 to AB084904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
XX The S. pneumoniae antigens have antibacterial activity and can be
XX used in vaccines. The S. pneumoniae antigens can also be used to
XX prevent or attenuate a Streptococcal infection in an animal. The
XX polynucleotides encoding the S. pneumoniae antigens can be used to
XX detect Streptococcus nucleic acids. AB084905 to AB085130 represent
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX which are used in an example from the present invention.
XX
XX Sequence 591 AA:

Query Match 24.7%; Score 55.5; DB 23; Length 591;
Best Local Similarity 27.1%; Pred. No. 30;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;
QY 3 YKNRYRYALKSGSVNAP-----MPENGQTEINDWVFMGYKQ 39
DB 190 FENGHHYYLKSGGYMANEMIMDKESWFLKFDGKMAEKEMWYDSSHQ 237

RESULT 13
AAV81517
ID AAV81517 standard; Protein; 659 AA.
XX
XX AAV81517;
XX
XX 24-MAY-2000 (first entry)
XX
XX Streptococcus pneumoniae type 4 protein sequence #17.
XX
XX Streptococcus pneumoniae: vaccine; screening; protein antigen;
XX antibacterial; antiinflammatory; meningitis; infection; diagnosis;
XX pneumococcal disease.
XX
XX Streptococcus pneumoniae.
XX
XX WO200006737-A2.
XX
XX 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02451.
XX
XX 27-JUL-1998; 98GB-0016337.
XX 19-MAR-1999; 99US-0125164.
XX
XX (MIGR-) MICROBIAL TECHINCS LTD.
XX
XX Gilbert CFG, Hansbro PM;
XX
XX WPI: 2000-195300/17.
XX
XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX or inhibiting expression of the protein -
XX
XX Claim 1; Page 74; 108pp; English.
XX
XX AAV81501 to AAV81679 represent specifically claimed protein sequences
XX
XX

CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 452 AA;

Query Match 24.4%; Score 55; DB 22; Length 452;
Best Local Similarity 38.5%; Pred. No. 26;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

QY . 12 KSGGSYNAPMPENG---QTENNDAV 33
I: | : : : |||
DB 200 KAKGEIFSGIPENGIAIMNADNDWL 225

Search completed: December 10, 2002, 10:54:04
Job time : 29.2 secs

US-09-741-669-450

Query Match 24.4%; Score 55; DB 10; Length 452;
Best Local Similarity 38.5%; Pred. No. 5.4;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY 12 KSGGSVNAMPENG---QTENNNDWV 33
1: | : : : ||||| : ||||| :
Db 200 KAKGEIFSGLPENGTAIMNADNDWL 225

RESULT 6

US-09-815-242-10027
Sequence 10027, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 10027
LENGTH: 452
TYPE: PRT
ORGANISM: Escherichia coli
US-09-815-242-10027

Query Match 24.4%; Score 55; DB 10; Length 452;
Best Local Similarity 38.5%; Pred. No. 5.4;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY 12 KSGGSVNAMPENG---QTENNNDWV 33
1: | : : : ||||| : ||||| :
Db 200 KAKGEIFSGLPENGTAIMNADNDWL 225

RESULT 7

US-09-815-242-14074
Sequence 14074, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard

TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 14074
LENGTH: 452
TYPE: PRT
ORGANISM: Salmonella typhi
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(452)
OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-14074

Query Match 24.4%; Score 55; DB 10; Length 452;
Best Local Similarity 38.5%; Pred. No. 5.4;
Matches 10; Conservative 5; Mismatches 7; Indels 4; Gaps 1;

OY 12 KSGGSVNAMPENG---QTENNNDWV 33
1: | : : : ||||| : ||||| :
Db 200 KAKGEITGLPENGTAIMNADNDWL 225

RESULT 8

US-09-815-242-5118
Sequence 5118, Application US/09815242
Patent No. US20020061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:54:49 ; Search time 9.8 Seconds
(without alignments)
268.490 Million cell updates/sec

Title: US-09-142-970-3_COPY_1_40
Perfect score: 225
Sequence: 1 LYXKRYVALKSGSVNAPMPENGQTEINNDWFMGYKOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 232864 seqs, 65780094 residues

Total number of hits satisfying chosen parameters: 232864

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	53.5	23.8	62	US-10-209-582-1019	Sequence 1019, Ap
2	50.5	22.4	451	US-09-724-676-69620	Sequence 69620, A
3	50.5	22.4	451	US-09-724-676A-69620	Sequence 69620, A
4	50.5	22.4	479	US-09-724-676-69617	Sequence 69617, A
5	50.5	22.4	479	US-09-724-676A-69617	Sequence 69617, A
6	50.5	22.4	518	US-09-724-676-69618	Sequence 69618, A
7	50.5	22.4	518	US-09-724-676-69619	Sequence 69619, A
8	50.5	22.4	518	US-09-724-676A-69618	Sequence 69618, A
9	50.5	22.4	518	US-09-724-676A-69619	Sequence 69619, A
10	50	22.2	132	US-09-724-676-76963	Sequence 76963, A
11	50	22.2	132	US-09-724-676A-76963	Sequence 76963, A
12	50	22.2	148	US-09-724-676-76966	Sequence 76966, A
13	50	22.2	148	US-09-724-676A-76966	Sequence 76966, A
14	49	21.8	417	PCT-US02-35143-14	Sequence 14, Appl
15	49	21.8	417	US-10-286-326-14	Sequence 14, Appl
16	48.5	21.6	378	US-10-216-209-87	Sequence 87, Appl
17	48	21.3	54	PCT-US02-32727-29179	Sequence 29179, A
18	48	21.3	54	US-10-057-498-29179	Sequence 29179, A
19	48	21.3	90	PCT-US02-32727-20352	Sequence 20352, A
20	48	21.3	90	US-10-057-498-20352	Sequence 20352, A
21	48	21.3	107	US-09-134-000C-4599	Sequence 4599, Ap
22	48	21.3	275	US-10-284-986-8	Sequence 8, Appl
23	48	21.3	275	US-10-285-042-8	Sequence 8, Appl
24	47.5	21.1	352	US-10-278-173-124	Sequence 124, App
25	47.5	21.1	352	US-10-278-536-172	Sequence 172, App
26	47.5	21.1	352	US-10-295-403-8	Sequence 8, Appl

27	47.5	21.1	439	US-09-134-000C-6557	Sequence 6557, Ap
28	47	20.9	104	US-09-513-999C-8076	Sequence 8076, Ap
29	47	20.9	208	US-09-724-676-93894	Sequence 93894, A
30	47	20.9	208	US-09-724-676A-93894	Sequence 93894, A
31	46.5	20.7	571	US-10-287-274-406	Sequence 406, Appl
32	46	20.4	378	US-10-274-694-10	Sequence 10, Appl
33	46	20.4	694	US-09-724-676-75372	Sequence 75372, A
34	46	20.4	694	US-09-724-676A-75372	Sequence 75372, A
35	46	20.4	792	US-09-724-676-75371	Sequence 75371, A
36	46	20.4	792	US-09-724-676A-75371	Sequence 75371, A
37	46	20.4	1044	US-09-724-676-88858	Sequence 88858, A
38	46	20.4	1044	US-09-724-676A-88858	Sequence 88858, A
39	46	20.4	1207	US-09-724-676-88853	Sequence 88853, A
40	46	20.4	1207	US-09-724-676A-88853	Sequence 88853, A
41	46	20.4	1235	US-09-724-676-88864	Sequence 88864, A
42	46	20.4	1235	US-09-724-676A-88864	Sequence 88864, A
43	46	20.4	1240	US-09-724-676-88862	Sequence 88862, A
44	46	20.4	1240	US-09-724-676A-88862	Sequence 88862, A
45	46	20.4	1269	US-09-724-676-88855	Sequence 88855, A

ALIGNMENTS

```

RESULT 1
US-10-209-582-1019
; Sequence 1019, Application US/10209582
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM009CIN
; CURRENT APPLICATION NUMBER: US/10/209,582
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 09/758,461
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1030
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1019
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-209-582-1019

Query Match      23.8%  Score 53.5:  DB 6:  Length 62:
Best Local Similarity 35.3%:  Pred. No. 2.5:
Matches 12:  Conservative 7:  Mismatches 14:  Indels 1:  Gaps 1:

DB      14  RGYTLRKGHTTASPDPESGVGRME-VFLGYSPF 46

RESULT 2
US-09-724-676-69620
; Sequence 69620, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: CompuGen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 CompuGen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28

```

```

; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 69620
; LENGTH: 451
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-724-676-69620

```

Query Match	22.48;	Score 50.5;	DB 5;	Length 451;
Best Local Similarity	31.78;	Pred. No. 50;		
Matches 13; Conservative	6;	Mismatches 3;	Indels 19;	Gaps 2;

```
QY      2 YKKNY--RYALKSGGSVNAPMENGOTENNMDVMFGKOE 400
        | ||: ||: ||: ||          |:|||::
DB      65 YIKNMRRPYFLKTGDS-----FIQKKEK 888
```

RESULT 3
US-09-724-676A-69620
; Sequence 69620, Application US/09724676A

```

1. APPLICANT: CompuGen LTD
2. TITLE OF INVENTION: Variants of alternative splicing
3. FILE REFERENCE: 129181.4 CompuGen
4. CURRENT APPLICATION NUMBER: US/09/724.676A
5. CURRENT FILING DATE: 2000-11-28
6. NUMBER OF SEQ ID NOS: 97222
7. SOFTWARE: seqin version 3.2
8. SEQ ID NO 69620
9. LENGTH: 451
10. TYPE: PRT
11. ORGANISM: Homo sapiens
12. US-09-724-676A-69620

```

Query Match	22.4%;	Score 50.5;	DB 5;	length 451;
Best Local Similarity	31.7%;	Pred. No. 50;		
Matches 13; Conservative	6;	Mismatches 3;	Indels 19;	Gaps 2

```

Oy      2 YKKNY--RYALKSGGSVNAPMENGQTEENNDDWYMGYKOE 40
        |  ||  ||  ||  ||  |  ||  ||  ||  ||  ||  ||
Db      65 YINWRPRYFLKTDGS-----FIGYKEX 88

```

RESULT 4
US-09-724-676-69617
; Sequence 69617, Application US/09724676

? APPLICANT: CompuGen LTD
 ? TITLE OF INVENTION: Variants of alternative splicing
 ? FILE REFERENCE: 129181.4 CompuGen
 ? CURRENT APPLICATION NUMBER: 05/09/724.676
 ? CURRENT FILING DATE: 2000-11-28
 ? NUMBER OF SEQ ID NOS: 97222
 ? SOFTWARE: PatentIn version 3.2
 ? SEQ ID NO 69617
 ? LENGTH: 479
 ? TYPE: PRT
 ? ORGANISM: Homo sapiens
 ? OS-09-724-676-69617

Query Match	22.48;	Score 50.5;	DB 5;	Length 479;
Best Local Similarity	31.7%;	Pred. No. 53;		
Matches 13;	Conservative 6;	Mismatches 3;	Indels 19;	Gaps 2

```
QY      2 YYYKNY--RYVALKSGGSYNAPMBENGQTENNDDVMEGYKQE 40
          | ||: ||: ||: ||: | :|||::
Db     18 YKANRPPIFLTKDGS-----FIGKEK 41
```

RESULT 5
US-09-724-676A-69617
; Sequence 69617, Application US/09724676A
; GENERAL INFORMATION:

```

? APPLICANT: CompuGen LTD
? TITLE OF INVENTION: Variants of alternative splicing
? FILE REFERENCE: 129181.4 CompuGen
? CURRENT APPLICATION NUMBER: US/09/724.676A
? CURRENT FILING DATE: 2000-11-28
? NUMBER OF SEQ ID NOS: 97222
? SOFTWARE: patentIn version 3.2
? SEQ ID NO 69617

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-69617

```

Query Match	22.48;	Score 50.5;	DB 5;	Length 479;
Best Local Similarity	31.78;	Pred. No. 53;		
Matches 13; Conservative	6;	Mismatches 3;	Indels 19;	Gaps 2

```
QY      2 YKNY--RYALSGGSYNAPMPENGOTENNDAVMGTYKOE 400
        | ||: ||: ||: ||          |:|:::
Db     18 YKNWRPRYFLKTDS-----FIGYEK 411
```

RESULT 6
US-09-724-676-69618
; Sequence 696618, Application US/09724676

```

? APPLICANT: CompuGen LTD
? TITLE OF INVENTION: Variants of alternative splicing
? FILE REFERENCE: 129181.4 CompuGen
? CURRENT APPLICATION NUMBER: US/09/724,676
? CURRENT FILING DATE: 2000-11-28
? NUMBER OF SEQ ID NOS: 97222
? SOFTWARE: PatentIn version 3.2
? SEQ ID NO 69618
? LENGTH: 518
? TYPE: PRT
? ORGANISM: Homo sapiens
? OS-09-724-676-69618

```

Query Match	22.4%;	Score 50.5;	DB 5;	Length 518;
Best Local Similarity	31.7%;	Pred. No. 58;		
Matches 13; Conservative	6;	Mismatches 3;	Indels 19;	Gaps 2

```
0Y      2 YKKNY--RYALKSGGSVNAPMENGOTENNDMVEMGKOE 400
        | ||: ||: ||: ||          |:||||:
Db      18 YKNNRRRYFLKTDS-----FISGKEK 411
```

RESULT 7
US-09-724-676-69619
; Sequence 69619, Application US/09724676

```

: APPLICANT: Comugen LTD
: TITLE OF INVENTION: Variants of alternative splicing
: FILE REFERENCE: 129181.4 Comugen
: CURRENT APPLICATION NUMBER: US/09/724,676
: CURRENT FILING DATE: 2000-11-28
: NUMBER OF SEQ. ID NOS: 97222
: SOFTWARE: PatentIn version 3.2
: SEQ. ID NO 69619
:
: LENGTH: 518
:
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-724-676-69619

```

Query Match	22.4%;	Score 50.5;	DB 5;	Length 518;
Best Local Similarity	31.7%;	Pred. No. 58;		
Matches 13;	Conservative 6;	Mismatches 3;	Indels 19;	Gaps 2

```
07      2 YXKNY--RYALKSGGSYNAPMBENGOTENNMDWFMGYKOE 4C  
          | ||| :||| :|||   |:|||:  
Db     18 YTKNWRPRFYFLKTGDS-----FVIGREK 4I
```


[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:50:48 ; Search time 11 Seconds
(without alignments)
349.580 Million cell updates/sec

Title: US-09-142-970-3_COPY_1_40
Perfect score: 225
Sequence: 1 LYKKNRYVALKSGGSYNAPMPENGOTENDWVFMGYKOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	203	90.2	1561	2	S61314
2	203	90.2	1773	2	A81937
3	199	88.4	1815	2	C81169
4	198	88.0	1532	2	A26039
5	94	41.8	1541	2	A37023
6	94	41.8	1694	2	H64106
7	94	41.8	1702	2	A41859
8	88	39.1	1545	2	B41859
9	86	38.2	1849	2	C41859
10	59	26.2	221	2	D90603
11	58.5	26.0	114	2	AD0785
12	58.5	26.0	532	1	CPBYX
13	57.5	25.6	293	2	C75421
14	56	24.9	1367	2	T33819
15	55.5	24.7	658	2	E95111
16	55.5	24.7	721	2	C97980
17	55	24.4	452	2	B90640
18	55	24.4	452	2	B85491
19	55	24.4	452	2	A10517
20	55	24.4	452	2	F64730
21	54.5	24.2	766	2	T20003
22	54.5	24.2	993	2	A38437
23	53.5	23.8	291	2	S03230
24	53.5	23.8	1317	2	F83310
25	53	23.6	152	2	H64749
26	53	23.6	596	2	T23193
27	53	23.6	646	2	T28868
28	53	23.6	821	2	C64461
29	53	23.6	2279	2	T42531

30	52.5	23.3	362	2	A12113	hypothetical prote
31	52.5	23.3	850	2	S20462	RNA12 protein - ye
32	52	23.1	343	2	G96533	hypothetical prote
33	52	23.1	350	2	S47292	phenol 2-monooxyge
34	52	23.1	389	2	B87001	probable acyl-CoA
35	52	23.1	471	2	H82814	replicative acyl-CoA
36	52	23.1	478	2	F90497	hypothetical prote
37	52	23.1	571	2	T26638	hypothetical prote
38	52	23.1	1356	2	T16718	hypothetical prote
39	52	23.1	2280	2	T38906	hypothetical prote
40	51.5	22.9	671	2	A38109	acetyl-CoA carboxy
41	51	22.7	515	2	S20493	autolysin - Entero
42	51	22.7	607	2	AB3511	endoglucanase - Cl
43	51	22.7	623	2	T33969	oligodeoxyribonuclease
44	51	22.7	641	2	A55549	hypothetical prote
45	51	22.7	1144	2	T21223	glucan 1,6-alpha-l

ALIGNMENTS

RESULT 1
S61314
IGA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HF13
C>Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61314
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; PMID:95302961; PMID:7783620
A:Accession: S61314
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1561 <LOW>
A:Cross-references: EMBL:X62474; NID:g732873; PIDN:CAA57857.1; PID:g732874
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 90.2% Score 203. DB 2; Length 1561;
Best Local Similarity 90.0%; Pred. No. 3.5e-18;
Matches 36; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

DB 584 LYKKNRYVALKSGGSYNAPMPENGOTENDWVFMGYKOE 40
|||||
1 LYKKNRYVALKSGGSYNAPMPENGOTENDWVFMGYKOE 40

RESULT 2
A81937
IGA-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [Imported] - Neisseria menin
N:Alternate names: Iga1 protease; Iga1 protease (EC 3.4.21.7) [misnomer]; Immunogl
C:Species: Neisseria meningitidis
A:Variety: strain 22491 serogroup A; strain HF117; strain HF159; strain SM1027
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churchev, C.; Klee, S.R.; Mo
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491
A:Reference number: A81775; PMID:20222556; PMID:10761919
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737
A:Experimental source: serogroup A, strain 22491
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in Neiss
A:Reference number: S61314; PMID:95302961; PMID:7783620
A:Accession: S61317
A:Status: nucleic acid sequence not shown

RESULT 6

H64106

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain

N:Alternate names: immunoglobulin A1 proteolase type 1

C:Species: Haemophilus influenzae

A:Variety: strain Rd KW20

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000

C:Accession: H64106; A41500

R:Flischman, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Cocayne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghagen, N.S.M.

Science 269, 436-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64106

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1694 <TRG>

A:Cross-references: GB:U32779; GB:L42023; NID:g1574009; PIDN:AA22651.1; PID:g1574019; T

A:Experimental source: strain Rd KW20

R:Grund, F.J.; Plant, A.G.; Wright, A.

Infect. Immun. 58, 320-331, 1990

A:Title: Localization of the cleavage site specificity determinant of Haemophilus influe

A:Reference number: A41500; MUID:90129281; PMID:2105270

A:Accession: A41500

A:Molecule type: DNA

A:Residues: 1-377 <GRU>

A:Cross-references: GB:X59800

A:Experimental source: strain Rd KW20

A:Note: the authors translated the codon TGG for residue 319 as Thr

C:Function:

A:Description: this proteinase is classified as type 1 because it cleaves at a proline-s

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 41.8%; Score 94; DB 2; Length 1694;

Best Local Similarity 41.7%; Pred. No. 0.00076;

Matches 15; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 1 LYKNTRYALKSGSVNAPMPENGQTEINDVFMG 36

DB 601 INLENTYALKRKGASTRELSPKNGSENNMLYMG 636

RESULT 7

A41859

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influe

C:Species: Haemophilus influenzae

A:Variety: strain HK715

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: A41859

R:Poulsen, K.; Reinholdt, J.; Killian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: A41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1702 <POU>

A:Cross-references: GB:M87489; NID:g148906; PIDN:AA24967.1; PID:g148907

A:Experimental source: strain HK715

A:Note: sequence extracted from NCBI backbone (NCBIP:97282)

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 41.8%; Score 94; DB 2; Length 1702;

Best Local Similarity 41.7%; Pred. No. 0.00077;

Matches 15; Conservative 8; Mismatches 13; Indels 0; Gaps 0;

QY 1 LYKNTRYALKSGSVNAPMPENGQTEINDVFMG 36

DB 601 INLENTYALKRKGASTRELSPKNGSENNMLYMG 636

RESULT 8

B41859

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK393

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: B41859

R:Poulsen, K.; Reinholdt, J.; Killian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: B41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1545 <POU>

A:Cross-references: GB:M87490; NID:g148908; PIDN:AA24967.1; PID:g148909

A:Experimental source: strain HK393

A:Note: sequence extracted from NCBI backbone (NCBIP:97283)

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 39.1%; Score 88; DB 2; Length 1545;

Best Local Similarity 42.1%; Pred. No. 0.0042;

Matches 16; Conservative 7; Mismatches 13; Indels 2; Gaps 1;

QY 1 LYX-KNRYRYALKSGSVNAPMPENGQTEINDVFMG 36

DB 597 LYFENRTRYALKKDGASTRELSPQNNGSENNMLYMG 634

RESULT 9

C41859

Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK613

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: C41859

R:Poulsen, K.; Reinholdt, J.; Killian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: C41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613

A:Note: sequence extracted from NCBI backbone (NCBIP:97285)

C:Superfamily: Iga-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 38.2%; Score 86; DB 2; Length 1849;

Best Local Similarity 39.5%; Pred. No. 0.0094;

Matches 15; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 1 LYX--NRYRYALKSGSVNAPMPENGQTEINDVFMG 36

DB 604 LYFNDNRSTYTLKKGASTRELSPQNSGSENNMLYMG 641

RESULT 10

D90603

Hypothetical protein MYPV_7320 [imported] - Mycoplasma pulmonis (strain UAB CTIP)

C:Species: Mycoplasma pulmonis

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 03-Aug-2001

C:Accession: D90603

R:Chenbaud, I.; Hellig, R.; Ferris, S.; Barbe, V.; Samson, D.; Galisson, F.; Moszer,

Nucleic Acids Res. 29, 2145-2153, 2001

A:Title: The complete genome sequence of the murine respiratory pathogen Mycoplasma p

A:Reference number: A99512; MUID:21267165; PMID:11353084

A:Accession: D90603

A:Status: preliminary

DB 143 YRFFAYYPVKRSTWVSGTGTTAPTASNDGDEPNDAVWVLAEXRK 188

RESULT 14

T33819

hypothetical protein W05F2.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T33819

R:Bradshaw, H.; Graves, T.; Blair, T.

Submitted to the EMBL Data Library, November 1998

A:Description: The sequence of C. elegans cosmid W05F2.

A:Reference number: 221418

A:Accession: T33819

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1367 <BNA>

A:Cross-references: EMBL:AF106582; PIDN:AMC78217.1; GSPDB:GN00019; CESP:W05F2.7

A:Experimental source: strain Bristol N2; clone W05F2

C:Genetics:

A:Gene: CESP:W05F2.7

A:Map position: 1

A:introns: 43/1; 106/1; 127/2; 161/1; 193/1; 279/1; 331/1; 393/1; 526/1; 548/1; 569/1; 6

Query Match

24.9% Score 56; DB 2; Length 1367;

Best Local Similarity 34.3% Pred. No. 58;

Matches: 12; Conservative 7; Mismatches 10; Indels 6; Gaps 1;

QY 6 YRYVALKSGSVNAPMPENGQTEINDWFMGYKQE 40

DB 150 YRYKISNGVFHA-----YSPDQVIVYDGYNQE 178

RESULT 15

E95111

endo-beta-N-acetylglucosaminidase [Imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001

C:Accession: E95111

R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid

on, J.D.; Unayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,

nson, T.; Hickey, E.K.; Holt, I.E.

Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: E95111

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-658 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75086.1; PID:q14972439; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

A:Gene: SP0965

Query Match

24.7% Score 55.5; DB 2; Length 658;

Best Local Similarity 27.1% Pred. No. 30;

Matches: 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

QY 3 YKNRYVALKSGSVNAP-----MPENGQTEINDWFMGYKQ 39

DB 213 FENGHYIYLLKSGVMAANEMWDKESWFLKFDGKMAEKEMVYDSHSQ 260

Search completed: December 10, 2002, 10:57:39
Job time: 13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:17:38 ; Search time 5.6 Seconds

(without alignments)
296.259 Million cell updates/sec

Title: US-09-142-970-3_COPY_1_40

Perfect score: 225

Sequence: 1 LYKKNRYALKSGSVNAPMPENGOTENNDAVFMGKGE 40

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt-40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES:

Result No.	Score	Query Match	Length	ID	Description
1	198	88.0	1532	1	IGA_NEIGO
2	94	41.8	1541	1	IGA1_HAEIN
3	94	41.8	1594	1	IGA0_HAEIN
4	94	41.8	1702	1	IGA2_HAEIN
5	88	38.2	1545	1	IGA3_HAEIN
6	86	36.2	1849	1	IGA4_HAEIN
7	58.5	26.0	532	1	CBPT_YEAST
8	55.5	24.7	658	1	LYTB_STRPN
9	55	24.4	452	1	MURP_ECOLI
10	54.5	23.8	993	1	TSH_DROME
11	53.5	22.2	291	1	Y32K_SSV1
12	53	23.6	152	1	YAPX_ECOLI
13	53	23.6	230	1	CLD2_CANFA
14	52.5	23.3	850	1	RNI2_YEAST
15	52	23.1	389	1	ACDP_MYCLE
16	52	23.1	2280	1	COAC_SCHPO
17	51.5	22.9	671	1	ALYS_ENTFA
18	51	22.7	452	1	EREL_PRYBB
19	51	22.7	515	1	GUND_CLOCL
20	51	22.7	641	1	IMD_ARGO
21	50.5	22.4	114	1	YEGJ_ECOLI
22	50.5	22.4	454	1	AKT3_RAT
23	50.5	22.4	479	1	AKT3_HUMAN
24	50.5	22.4	479	1	AKT3_MOUSE
25	50	22.2	295	1	RP32_CAUCR
26	50	22.2	334	1	DPOB_HUMAN
27	50	22.2	334	1	DPOB_RAT
28	50	22.2	359	1	CD72_HUMAN
29	50	22.2	2334	1	WAPA_BACSU
30	49.5	22.0	507	1	YME3_YEAST
31	49.5	22.0	874	1	SYV_STRCO
32	49	21.8	297	1	YMY9_YEAST
33	49	21.8	318	1	IKBA_CHICK

34	49	21.8	417	1	SOXC_RHOSO
35	49	21.8	512	1	SING_DROME
36	49	21.8	954	1	XVNA_RUMFL
37	48.5	21.6	339	1	NCF4_MOUSE
38	48.5	21.6	378	1	RFBP_SALBO
39	48.5	21.6	510	1	CEB1_PIG
40	48.5	21.6	1032	1	VG07_BPT4
41	48.5	21.6	1729	1	TABP_HUMAN
42	48	21.3	66	1	TXMA_DENPO
43	48	21.3	100	1	YF78_MYCPN
44	48	21.3	269	1	ESL1_MYCPN
45	48	21.3	327	1	ACCO_DORSP

ALIGNMENTS

RESULT 1
IGA_NEIGO STANDARD: PRT: 1532 AA.
AC P09790;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI-TAXID:485;
RN [1]
RP SEQUENCE FROM N.B., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RX MEDLINE=87115823; PubMed=3027577;
RA Pohner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease.";
RL J. Biol. Chem. 265:3738-3743(1990).
RN Nature 325:458-462(1987).
RL [2]
RP ACTIVE SITE.
RX MEDLINE=9015452; PubMed=2105953;
RA Bachovchin W.W., Platt A.G., Plentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of IGA1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids.";
RL J. Biol. Chem. 265:3738-3743(1990).
CC -!- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -!- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X04835; CAA28538.1; -;
CC PIR: A26039; A26039.
CC MEROPS: S06.001; -;
CC InterPro: IPR000710; IGA_S6.
CC InterPro: IPR004899; Peptidase_sup.
CC Pfam: PF02395; IGA1; 1.
CC Pfam: PF03212; Peptidase_1.
CC PRINTS: PR00921; IGASERPTASE.
CC Hydrolase; Serine protease; Zymogen; Autocatalytic cleavage;


```

CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL; X59800; -- NOT_ANNOTATED_CDS.
DR EMBL; U32779; AAC22651.1; --
DR MEROPS; S06.001; --
DR TIGR; H10990; --
DR InterPro; IPR000710; IGA_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; IGAL; 1.
DR Pfam; PF03212; Pertactin; 2.
DR PRINTS; PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 25
FT CHAIN 26 1014
FT PROPEP 1015 1694
FT ACT_SITE 288 288
FT CONFLICT 253 254
FT CONFLICT 272 272
FT CONFLICT 464 464
FT CONFLICT 866 866
FT CONFLICT 1036 1036
FT CONFLICT 1074 1074
FT CONFLICT 1421 1421
FT CONFLICT 1545 1545
SO SEQUENCE 1694 AA; 185539 MW; C52427013F93178C CRC64;
POTENTIAL.
IMMUNOGLOBULIN A1 PROTEASE.
HELPER PEPTIDE (POTENTIAL).
PROBABLE.
EN -> GV (IN REF. 1).
G -> A (IN REF. 1).
G -> E (IN REF. 1).
S -> T (IN REF. 1).
A -> D (IN REF. 1).
A -> G (IN REF. 1).
A -> G (IN REF. 1).
H -> T (IN REF. 1).

```

CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL: M87489; AAAA24966.1; -.
DR	MEROPS; S06.001;-.-.
DR	InterPro; IPR000710; IGA_S6.
DR	InterPro; IPR004899; Pertact_sup.
DR	Pfam; PF03395; IGA1; 1.
DR	Pfam; PF03212; Pertactin; 2.
.DR	PRINTS; PRO0921; ICASEKPTASE.
KM	Hydrolase; Serine protease; Transmembrane; Zymogen; Repeat; Signal.
FT	SIGNAL 1 25
FT	CHAIN 26 1014
FT	PROPEP 1015 1702
FT	ACT_SITE 288 288
FT	DOMAIN 1109 1124
FT	
FT	REPEAT 1109 1116
FT	REPEAT 1117 1124
FT	
SEQUENCE	1702 AA; 186539 MW; 860F70D2667807A6 CRC64;
Query Match	41.8%; Score 94; DB 1; Length 1702;
Best Local Similarity	41.7%; Pred. No. 0.00029;
Matches 15; Conservative	8; Mismatches 13; Indels 0; Gaps 0;

Query Match	41.8%	Score 94	DB 1	Length 1694
Best Local Similarity	41.7%	Pred. No. 0.00029		
Matches 15, Conservative		8, Mismatches 13,	Indels 0,	Gaps 0

QY	1	LYKKNRYTALKSGGSVNA	PM	PENGQTENDNDWY	FMG	36
Db	601	LNLENYYTALKRGASTRESELP <td>PKNSGSE <td>NENWYMG</td> <td></td> <td>636</td> </td>	PKNSGSE <td>NENWYMG</td> <td></td> <td>636</td>	NENWYMG		636

```

Qy      1 LYYKNRYALKSGGSYNAMPPEQGOTENNDDWVEMG 36
          |::|||::|::|::|::|::|
Db      601 LNLNRYTVALRKGASTRSELPRNSGESNENWLYMG 636

```

RESULT 5
IGAS_HAEIN

RESULT 4	IGA2_HAEIN	STANDARD;	PRT; 1702 AA.
ID	IGA2_HAEIN		
AC	P45384;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease)		

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGAL protease)
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales; Pasteurellaceae; Haemophilus.

Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

RP SEQUENCE FROM N.A.

OX NCBI_TaxID=727;

RX MEDLINE=9223494

RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;

RT "A comparative genetic study of serologically distinct *Haemophilus*
RT *influenzae* type 1 immunoglobulin A1 proteases.",
RL J. Bacteriol. 174:2913-2921(1992).

RT "A comparative genetic study of serologically distinct *Haemophilus*
RT influenzae type 1 immunoglobulin A1 proteases.",
RL J. Bacteriol. 174:2913-2921(1992).

CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC

CC - FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.

CC substrates are known.
CC -I- SUBCELLULAR LOCATION: Secreted.

CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at certain Pro-1-Xaa bonds in the hinge region. No small molecule substrates are known.

CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PEPTIDOLYTIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE

CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE AND THE CADOXY-TERMINAL HELPER DOMAIN ACCOMPLISHES HYDROPHOBIC
CC

CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).

CC OUTER MEMBRANE TO FORM A PORE FOR CRECRETION OF THE PROTEASE
CC DOMAIN THE HELPER DOMAIN IS THEN RELEASED BY ATPOTROTEOLYSTS /BY
CC SITE), AND THE CARBOAL IDENTICAL HELPER DOMAIN ASSOCIATES WITH THE

[illegible]

CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6
CC SIMILARITY).

the European Bioinformatics Institute. There are no restrictions on its use. This manuscript is copyright, it is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - CC BY-NC-ND 4.0 International license.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the Swiss Institute of Bioinformatics.

use by non-profit institutions as long as its content is not modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sih.ch/announcements>).

RT "2.8-A structure of yeast serine carboxypeptidase.";
 RL Biochemistry 33:1106-1120(1994).
 CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS
 CC PRESENTLY IN PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC
 CC RESIDUE IN P1' POSITION, AS WELL AS METHIONINE, LEUCINE OR
 CC PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
 CC broad specificity.
 CC -1- ENZYME REGULATION: INHIBITED BY ZPPK.
 CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
 CC -1- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZYMOGEN AND
 CC IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO
 CC A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE
 CC GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE
 CC GOLGI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
 CC SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS
 CC SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR
 CC ENDOPLASM-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
 CC N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY
 CC ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 10.
 CC -1- DATABASE: NAME=Northington enzyme manual;
 CC WWW="http://www.worthington-biochem.com/manual/C/COY.html".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M15482; AAA34902.1;
 DR EMBL: X80836; CAA56806.1;
 DR PIR: A26597; CPBY.
 DR PDB: 1YSC; 22-JUN-94.
 DR PDB: 1CPY; 15-SEP-95.
 DR MEROPS: S10.001;
 DR SGD: S0004912; PC1.
 DR InterPro: IPR000379; Ser_estra.site.
 DR InterPro: IPR001563; Serine_catpept.
 DR Pfam: PF00450; Serine_catpept.1.
 DR PRINTS: PR00724; CROXYPTASEC.
 DR PRODOM: PD001189; Serine_catpept.1.
 DR PROSITE: PS00131; CARBOXYPEPT_SER_SER.1.
 DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS.1.
 KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal;
 KW 3d-structure.
 FT SIGNAL 1 20
 FT PROPEP 21 111
 FT CHAIN 112 532
 FT ACT_SITE 257 257
 FT ACT_SITE 449 449
 FT ACT_SITE 508 508
 FT BINDING 452 452
 FT BINDING 509 509
 FT DISULFID 167 409
 FT DISULFID 304 318
 FT DISULFID 328 351
 FT DISULFID 335 344
 FT DISULFID 373 379
 FT CARBOHYD 124 124
 FT CARBOHYD 198 198
 FT CARBOHYD 279 279
 FT CARBOHYD 479 479
 FT SITE 24 27
 FT MUTAGEN 508 508
 FT CONFLICT 260 261
 FT CONFLICT 389 389
 FT CONFLICT 529 529
 FT SEQUENCE 532 AA; 59802 MW; 7227F3489CDBD952 CRC64;
 Query Match 26.0%; Score 58.5; DB 1; Length 532;

Best Local Similarity 33.3%; Pred. No. 3.7;
 Matches 12; Conservative 9; Mismatches 14; Indels 1; Gaps 1;
 Oy 3 YKNRYRYALKSGG-SYNAPMPENGOTENDWVPMGY 37
 Db 495 YKHYLRVFNQNGHWPEDVPENALSMVNEWTHGCF 530
 RESULT 8
 ID LYTEB_STREP STANDARD: PRT; 658 AA.
 AC Q924P7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.96)
 DE (Murein hydrolase).
 GN LYTEB OR SP0965.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
 RC STRAIN=R6;
 RX MEDLINE=99195827; PubMed=10096093;
 RA Garcia P., Gonzalez M.P., Garcia E., Lopez R., Garcia J.L.;
 RT "LyteB, a novel pneumococcal murein hydrolase essential for cell
 RT separation.";
 RL Mol. Microbiol. 31:1275-1281(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TRG4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 RA Holtzapple E., Khouri H., Wolf A.M., Ueberback T.R., Hansen C.L.,
 RA McDonald L.A., Feldblyum T.V., Angiloul S., Dickinson T., Hickey E.K.,
 RA Dohlt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
 RT "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
 CC CELL SEPARATION.
 CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
 CC unit in high-mannose glycopeptides and glycoproteins containing
 CC the -(Man(GlcNAc)2)Asn-structure. One N-acetyl-D-glucosamine
 CC residue remains attached to the protein; the rest of the
 CC oligosaccharide is released intact.
 CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AJ010312; CAA09078.1;
 DR EMBL: AE007400; AAK75086.1;
 DR TIGR: SP0965;
 DR InterPro: IPR002901; Amidase_4.
 DR Pfam: PF01832; Amidase_4; 1.
 DR SMART: SM00047; LY22; 1.
 DR Signal: Hydrolase; Cell wall; Complete proteome.
 FT SIGNAL 1 23
 FT CHAIN 24 658
 FT CONFLICT 336 336
 FT PUTATIVE ENDO-BETA-N-
 FT ACETYLGUCOSAMINIDASE.
 FT I -> M (IN REF. 1).

```

FT CONFLICT 381 381 T -> A (IN REF. 1).
FT CONFLICT 384 384 E -> K (IN REF. 1).
FT CONFLICT 535 535 L -> P (IN REF. 1).
FT CONFLICT 580 580 F -> S (IN REF. 1).
SQ SEQUENCE 658 AA; 76469 MW; B625515006C9CB36 CRC64;

Query Match
Best Local Similarity 24.7%; Score 55.5; DB 1; Length 658;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

OY 3 YKNRYRYALKSGSVNAP-----MPENGOTENNDFWFMCKYQ 39
DB 213 FENGHYIYLLKSGGYMANEMIMDKSEWFLKFDGKMAEKEMWYDSHSQ 260

RESULT 9
MURF_ECOLI STANDARD; PRT: 452 AA.
AC P11880; P77636; O07100;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylmuramoyl-alanyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAc-pentapeptide synthetase)
GN MURF OR MRA OR B0086.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=89345095; PubMed=2668880;
RA Parquet C., Flouret B., Mengin-Jecreulx D., van Heijenoort J.;
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNAc-
RT pentapeptide synthetase of Escherichia coli.";
RL Nucleic Acids Res. 17:5379-5379(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=92334977; PubMed=1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MB2884;
RX MEDLINE=97128642; PubMed=8973200;
RA Anderson M.S., Eveland S.S., Onishi H.R., Pompliano D.L.;
RT "Kinetic mechanism of the Escherichia coli UDPMurNAc-tripeptide
RT D-alanyl-D-alanine-adding enzyme: use of a glutathione S-transferase
RT fusion.";
RL Biochemistry 35:16264-16269(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A., AND MUTANT MURF2.
RC STRAIN-CGSC 5990;
RX MEDLINE=97309380; PubMed=9166795;
RA Eveland S.S., Pompliano D.L., Anderson M.S.;
RT "Conditionally lethal Escherichia coli murein mutants contain point
RT defects that map to regions conserved among murein and Tolyl poly-
```

```

RT gamma-glutamate ligases: identification of a ligase superfamily.";
RL Biochemistry 36:6223-6229(1997).
RN [6]
RP CHARACTERIZATION, AND SEQUENCE OF 1-15.
RX MEDLINE=90248455; PubMed=2186811;
RA Duncan K., van Heijenoort J., Walsh C.T.;
RT "Purification and characterization of the D-alanyl-D-alanine-adding
RT enzyme from Escherichia coli.";
RL Biochemistry 29:2379-2386(1990).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=20545602; PubMed=11090285;
RA Yan Y., Munshi S., Leitling B., Anderson M.S., Chrzas J., Chen Z.;
RT "Crystal structure of Escherichia coli UDPMurNAc-tripeptide
RT D-alanyl-D-alanine-adding enzyme (MurF) at 2.3-A resolution.";
RL J. Mol. Biol. 304:435-445(2000).
CC -1- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN.
CC -1- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -1- PATHWAY: peptidoglycan biosynthesis.
CC -1- SUBUNIT: MONOMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE MURCDF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
DR EMBL: X15432; CAA34473.1; -
DR EMBL: X55034; CAA38863.1; -
DR EMBL: D10483; BAA01351.1; -
DR EMBL: U67891; AAC44657.1; -
DR EMBL: AE000118; AAC73197.1; -
DR EMBL: U67893; AAB60788.1; -
DR PIR: S04846; S04846.
DR PIR: S40596; S40596.
DR PDB: 1GC4; 20-DEC-00.
DR EcoGene: EG10622; murF.
DR InterPro: IPR000713; Mur_ligase.
DR InterPro: IPR004101; Mur_ligase_C.
DR Pfam: PF01225; Mur_ligase; 1.
DR TIGRFAMs: TIGR01143; murF; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW ATP-binding; 3D-structure; Complete proteome.
FT NP_BIND 107 113 ATP (POTENTIAL).
FT VARIANT 288 288 A -> T (IN MURF2; TS MUTANT WITH LOW
FT ACTIVITY).
FT CONFLICT 61 61 G -> A (IN REF. 1 AND 2).
FT CONFLICT 178 178 A -> R (IN REF. 1 AND 2).
SQ SEQUENCE 452 AA; 47447 MW; B46E2E57BDBBC572 CRC64;

Query Match
Best Local Similarity 24.4%; Score 55; DB 1; Length 452;
Matches 10; Conservative 6; Mismatches 6; Indels 4; Gaps 1;

OY 12 KSGSVNAPMPENG-----QFENNDFV 33
DB 200 KAKGEIFSGPENGIAIMNADNDWL 225

RESULT 10
TSH_DROME STANDARD; PRT: 993 AA.
ID TSH_DROME
AC P22265;
```

DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Teashirt protein.
 GN TSH.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 NCBI_TaxID=7227;
 OX
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91098655; PubMed=1846092;
 RA Fasano L., Roeder L., Core N., Alexandre E., Voia C., Jacq B.,
 RA Kerridge S.;
 RT "The gene teashirt is required for the development of Drosophila
 RT embryonic trunk segments and encodes a protein with widely spaced
 RT zinc finger motifs."
 RL Cell 64:63-79(1991).
 RN
 RP
 RP POSSIBLE FUNCTION.
 RX MEDLINE=93083418; PubMed=1360402;
 RA Roeder L., Voia C., Kerridge S.;
 RT "The role of the teashirt gene in trunk segmental identity in
 RT Drosophila."
 RL Development 115:1017-1033(1992).
 RN
 RN
 RP
 RP POSSIBLE FUNCTION.
 RX MEDLINE=95009555; PubMed=7925029;
 RA de Zulueta P., Alexandre E., Jacq B., Kerridge S.;
 RT "Homeotic complex and teashirt genes co-operate to establish trunk
 RT segmental identities in Drosophila."
 RL Development 120:2287-2296(1994).
 CC
 CC -I- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF DROSOPHILA EMBRYONIC
 CC TRUNK SEGMENTS. REPRESS THE EXPRESSION OF HEAD HOMOTIC GENES.
 CC NECESSARY, IN COMBINATION WITH SCR, FOR THE FORMATION OF THE
 CC PROHOMERIC SEGMENT. MAY BE INVOLVED IN CHROMATIN STRUCTURE FOR
 CC MODULATION OF TRANSCRIPTION. NEGATIVELY CONTROL THE EXPRESSION OF
 CC MOD AND POSITIVELY THAT OF DLX AND OF ITS OWN EXPRESSION.
 CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE TRUNK (FROM PS3
 CC TO PS13).
 CC
 CC -I- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL AND
 CC ADULT DEVELOPMENT. NOT MATERNALLY EXPRESSED.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; M57496; AAA28983.1; -
 CC PIR; A38437; A38437.
 CC TRANSFAC; T00805; -
 CC DR Flybase; FBgn0003866; tsh.
 CC DR InterPro; IPR000822; znf.C2H2.
 CC DR Pfam; PR00096; zf-C2H2.3.
 CC DR SMART; SM00355; znf-C2H2.3.
 CC DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 CC DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
 CC DR Developmental Protein; Transcription regulation; Repressor; Activator;
 CC Zinc-finger; Metal-binding; Repeat; DNA-binding; Nuclear protein.
 KW
 FT DOMAIN 104 136
 FT 175 183 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 354 557
 FT 378 378 ZINC FINGERS.
 FT ZN_FING 354 557 C2H2-TYPE.
 FT ZN_FING 466 490 C2H2-TYPE.
 FT ZN_FING 533 557 C2H2-TYPE.
 FT DOMAIN 104 107 POLY-ALA.
 FT 115 122 POLY-ALA.
 FT DOMAIN 175 180 POLY-GLU.

FT DOMAIN 401 407 POLY-PRO.
 FT DOMAIN 830 834 POLY-ASN.
 SQ SEQUENCE 993 AA; 106206 MW; 2DF9C6774F68B6D1 CRC64;
 Query Match 24.2%; Score 54.5; DB 1; Length 993;
 Best Local Similarity 35.3%; Pred. No. 25;
 Matches 12; Conservative 6; Mismatches 7; Indels 9; Gaps 1;
 QY 2 YKKNRYVALKSGS-----VNAPPENQ 26
 DB 740 YVQHYRTSSENGSECSAEPRLDAPTPKEQ 773
 RESULT 11
 Y32K_SSV1
 ID Y32K_SSV1 STANDARD; PRT; 291 AA;
 AC P20197;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Hypothetical 31.7 kDa protein (ORF A-291).
 OS Sulfolobus virus-like particle SSV1.
 OC Sulfolobus virus-like particle SSV1.
 OC Viruses; dsDNA viruses, no RNA stage; Fuselloviridae; Fusellovirus.
 OX NCBI_TaxID=10476;
 RN
 RN
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92024080; PubMed=1926776;
 RA Palm P., Schleper C., Gramp B., Yeats S., McWilliam P., Reiter W.-D.,
 RA Zillig W.;
 RT "Complete nucleotide sequence of the virus SSV1 of the
 RT archaeobacterium Sulfolobus shibatae."
 RL Virology 185:242-250(1991).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; X07234; CAA30198.1; -
 CC PIR; S03230; S03230.
 CC KW Hypothetical protein.
 CC
 CC SEQUENCE 291 AA; 31669 MW; 84D231AB748DBF2A CRC64;
 SQ
 Query Match 23.8%; Score 53.5; DB 1; Length 291;
 Best Local Similarity 24.6%; Pred. No. 8; 7;
 Matches 14; Conservative 8; Mismatches 14; Indels 21; Gaps 2;
 QY 1 LYKKNRYVALKSGS-----SYNAP--MPENQETNNMVFENG 36
 DB 35 IYNNYNYSLNAEGFGFSEFNNSNNWVETNFISTINLPSSLPNNYQINNAYSIYVG 91
 RESULT 12
 YAFX_ECOLI
 ID YAFX_ECOLI STANDARD; PRT; 152 AA.
 AC P75676; P71286; Q9R2D9;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein yafx.
 GN YAFX OR B0248.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,


```

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mao B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN
RP
RC SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA Mizuno T., Makino K., Nakata A., Yura T., Sempel G., Mizobuchi K.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of the
RT 4.0 - 6.0 min (188,987 - 281,416bp) region.";
RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lashari D., Lew H., Lin D., Namath A., Oelner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SIMILARITY: STRONG, TO E.COLI VFUX.
CC -1- SIMILARITY: TO E.COLI PLASMIDS ANTIRESTRICTION PROTEIN KLC4/KLIC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE00133; AAC73351.1; -
DR EMBL; D83536; BAA77917.1; ALT_INIT.
DR EMBL; U70214; AAB08668.1; ALT_INIT.
DR EcoGene; EG13336; YafX.
DR InterPro: IPR004914; Antirestrict.
DR Pfam; PF03230; Antirestrict; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 152 AA; 17419 MW; C82749CB8C2C1731 CRC64;

Query Match 23.6%; Score 53; DB 1; Length 152;
Best Local Similarity 36.4%; Pred. No. 5;
Matches 8; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

Qy 6 YRRYALKSGSVNAPMPENGQT 27
Db 66 WSFYTLNGGAFMSPEPDNDT 87

RESULT 13
CILD2_CANFA STANDARD; PRT; 230 AA.
ID CILD2_CANFA
AC 095KM6;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Claudin-2.
GN CLDN2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RC MEDLINE=21206012; PubMed=11309408;
RA Furuse M., Furuse K., Sasaki H., Tsukita S.;
RT "Conversion of zonulae occludentes from tight to leaky strand type by
RT introducing claudin-2 into Madin-Darby canine kidney I cells.";
RL J. Cell Biol. 153:263-272(2001).
CC -1- FUNCTION: COMPONENT OF TIGHT JUNCTION (TJ) STRANDS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.

```

```

CC -1- SIMILARITY: BELONGS TO THE CLAUDIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF358907; AAK51433.1; -
DR InterPro: IPR001832; Claudin.
DR InterPro: IPR004031; PMP22_Claudin.
DR Pfam; PF00822; PMP22_Claudin; 1.
DR PROSITE; PS01346; CLAUDIN; 1.
KW Tight junction; Transmembrane.
FT TRANSMEM 8 28
FT TRANSMEM 82 102 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 163 183 POTENTIAL.
SQ SEQUENCE 230 AA; 24502 MW; 91B71C1E5CDCABE9 CRC64;

Query Match 23.6%; Score 53; DB 1; Length 230;
Best Local Similarity 30.6%; Pred. No. 7.8;
Matches 11; Conservative 6; Mismatches 19; Indels 0; Gaps 0;

Qy 2 YKRYRYALKSGSVNAPMPENGQTEINDVFMGY 37
Db 194 YDSDVQAOPLATRSPPRQGPRAKSFNSYSLTG 229

RESULT 14
RN12_YEAST STANDARD; PRT; 850 AA.
ID RN12_YEAST
AC P32843;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RNA12 protein.
DE RNA12 OR PRP12 OR YMR302C OR YMR952.OAC.
GN RNA12 OR PRP12 OR YMR302C OR YMR952.OAC.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28583 / FL100;
RX MEDLINE=92212295; PubMed=1557037;
RX Liang S., Alksne L., Warner J.R., Lacroste F.;
RT "RNA12+, a gene of Saccharomyces cerevisiae involved in pre-rRNA
RT maturation. Characterization of a temperature-sensitive mutant,
RT cloning and sequencing of the gene.";
RL Mol. Gen. Genet. 232:304-312(1992).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RT Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: PLAYS A DISPENSABLE ROLE IN EARLY MATURATION OF
CC PRE-RRNA. IT MAY PROVIDE A CORCTOR WHICH PLAYS A MINOR AND
CC DISPENSABLE ROLE IN PRE-RRNA MATURATION. THIS PROTEIN MIGHT
CC MIGRATE PROGRESSIVELY ALONG EACH PRE-RNA MOLECULE IN THE EARLY
CC STEPS OF THE RNA MATURATION PATHWAY AND BE RELEASED ONCE THE
CC MODIFICATION IT CATALYSES HAD BEEN COMPLETED.
CC -1- SIMILARITY: TO S.POMBE SPC83.05.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:14:59 ; Search time 28.2 Seconds

(without alignments)
189,008 Million cell updates/sec

Title: US-09-142-970-4_COPY_1_40

Perfect score: 224
Sequence: 1 LYKNRYRYALKSGRLNAPMPENGVANNDVWFMGTQE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 segs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	100.0	104	19	AAW61605
2	223	99.6	104	19	AAW61606
3	211	94.2	104	19	AAW61603
4	199	88.8	104	19	AAW61604
5	189	84.4	105	19	AAW65656
6	189	84.4	105	19	AAW61602
7	92	41.1	1541	11	AAW07304
8	63	28.1	834	20	AAW34542
9	63	28.1	907	20	AAW34408
10	61.5	27.5	971	21	AAW95687

11	61	27.2	452	22	AAU34434
12	61	27.2	452	22	AAU38481
13	61	27.2	452	22	AAW98402
14	58	25.9	3084	19	AAW50891
15	58	25.9	3084	22	AAE11215
16	56	25.0	1216	22	ABG09754
17	56	25.0	2901	22	ABG09763
18	56	25.0	3075	19	AAW50892
19	55.5	24.8	948	22	ABW58055
20	54.5	24.3	456	17	AAW98365
21	54.5	24.3	459	22	ABW71664
22	54.5	24.3	569	19	AAW61214
23	54.5	24.3	569	23	ABW54633
24	54.5	24.3	591	19	AAW55099
25	54.5	24.3	591	23	ABW54593
26	54.5	24.3	659	21	AAW81517
27	54.5	24.3	678	21	AAW81667
28	53.5	23.9	194	23	ABW07412
29	53.5	23.9	265	18	AAW43003
30	53.5	23.9	265	21	AAW51810
31	53.5	23.9	310	18	AAW43004
32	53.5	23.9	310	21	AAW51809
33	53.5	23.9	365	18	AAW43005
34	53.5	23.9	365	21	AAW51808
35	53.5	23.9	404	18	AAW43006
36	53.5	23.9	404	21	AAW51807
37	53.5	23.9	411	18	AAW43007
38	53.5	23.9	411	21	AAW51806
39	53.5	23.9	417	18	AAW43008
40	53.5	23.9	417	21	AAW51805
41	53.5	23.9	425	18	AAW43009
42	53.5	23.9	430	21	AAW51804
43	53.5	23.9	463	18	AAW43010
44	53.5	23.9	463	21	AAW51803
45	53.5	23.9	523	18	AAW43011

ALIGNMENTS

RESULT 1
AAW61605
ID AAW61605 standard; peptide: 104 AA.
XX
AC AAW61605;
XX
DT 27-OCT-1998 (first entry)
XX
DE Neisseria IgA1 protease fragment 4.
XX
KW Immunoglobulin protease; carrier: paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
OS Neisseria sp.
XX
PN WO9831791-A1.
XX
PD 23-JUL-1998.
XX
PF 20-JAN-1998; 98WO-EP00294.
XX
PR 21-JAN-1997; 97EP-0100883.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
PI Achtmann M, Moreau M;
DR WPI: 1998-414092/35.
PT New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and


```

DE Neisseria IgA1 protease fragment 3.
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX WO9831791-A1.
PN
XX 23-JUL-1998.
PD
XX 20-JAN-1998; 98WO-EP00294.
PF
XX 21-JAN-1997; 97EP-0100883.
PR
XX
XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
PI
XX WPI: 1998-414092/35.
DR
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 2; Fig 1; 32pp: English.
PS
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 104 AA:
SQ
Query Match 88.8%; Score 199; DB 19; Length 104;
Best Local Similarity 87.5%; Pred. No. 1.8e-19;
Matches 35; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
OY 1 LYYKNRYRYALKSGGRLNAPMPENGVAENNNDVFMGYTOE 40
DB 1 LYYKNRYRYALKSGGSVNAPEENGCTENNNDVFMGYKOE 40
RESULT 5
AAW65656
ID AAW65656 standard; peptide; 105 AA.
XX
XX AAW65656;
AC
XX 15-OCT-1998 (first entry)
DT
XX
XX 105-mer peptide used in polysaccharide-peptide conjugate.
DE
XX Polysaccharide-peptide conjugate; vaccine; linker moiety; adjuvant;
KM immune response.
XX
XX Synthetic.
OS
XX Key Location/Qualifiers
FT Modified-site 1 /note="N-terminal acetyl"
FT
XX
XX WO9831393-A2.
PN
XX 23-JUL-1998.
PD
XX 21-JAN-1998; 98WO-EP00654.
PF

```

```

XX
XX 21-JAN-1997; 97EP-0100884.
PR
XX
XX (INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA
XX
XX Mistretta N, Moreau M;
PI
XX
XX WPI: 1998-413820/35.
DR
XX
XX Polysaccharide-peptide conjugates, useful as, e.g. vaccines -
PT comprise peptide moiety with at least six amino acid residues,
PT polysaccharide chain with at least four repeat units, and linker
PT moiety
XX
XX Example 1; Page 14; 28pp: English.
PS
XX
XX The invention relates to: (A) polysaccharide-peptide conjugate (in which
CC the polysaccharide is immunogenic), comprising: (a) a peptide moiety
CC which has at least 6 amino acid residues, at least 1 of which is a
CC cysteine residue; (b) a polysaccharide chain comprising at least 4 repeat
CC units, and (c) a linker moiety bound to the thiol group of the cysteine.
CC The linker is also bound to: (i) native amino, hydroxyl or carboxyl
CC groups of the polysaccharide chain; (ii) amino groups created by
CC hydrolysis of native N-acetyl groups of the polysaccharide chain, or (iii)
CC functional groups introduced on the polysaccharide chain upon
CC derivatisation with a spacer moiety bound to native amino, hydroxyl or
CC carboxyl groups of the polysaccharide chain, and (b) conjugating a
CC peptide as in (Aa) to a polysaccharide chain comprising at least 4 repeat
CC units, comprising: (a) coupling the peptide to a linker through the thiol
CC group of the cysteine residue, and (b) coupling the linker to the
CC polysaccharide chain through one of groups (i), (ii) or (iii) as
CC described in (A). The conjugates are especially useful as vaccines to
CC elicit a protective long term immune response against a pathogenic
CC microorganism from which the immunogenic polysaccharide chain is derived.
CC known polysaccharide-peptide conjugates are less immunogenic than their
CC polysaccharide-protein counterparts and require adjuvantation. The new
CC conjugates have good immunogenicity, largely because the peptides are of
CC sufficient size. The present sequence represents a peptide used in
CC a polysaccharide-peptide conjugate.
XX
XX Sequence 105 AA:
SQ
Query Match 84.4%; Score 189; DB 19; Length 105;
Best Local Similarity 82.5%; Pred. No. 4.2e-18;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
OY 1 LYYKNRYRYALKSGGRLNAPMPENGVAENNNDVFMGYTOE 40
DB 2 LYYKNRYRYALKSGGSVNAPEENGCTENNNDWILMGSTOE 41
RESULT 6
AAW61602
ID AAW61602 standard; peptide; 105 AA.
XX
XX AAW61602;
AC
XX 27-OCT-1998 (first entry)
DT
XX
XX Neisseria IgA1 protease fragment 1.
DE
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
OS
XX WO9831791-A1.
PN
XX 23-JUL-1998.
PD
XX 20-JAN-1998; 98WO-EP00294.
PF
XX 21-JAN-1997; 97EP-0100883.
PR

```

```

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INNR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
XX
XX WPI: 1998-414092/35.
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 6; Page 10; 32pp; English.
XX
XX The Neisseria immunoglobulin protease fragments AAM61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 105 AA;
SQ
XX
XX Query Match 84.4%; Score 189; DB 19; Length 105;
XX Best Local Similarity 82.5%; Pred. No. 4,2e-18;
XX Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
OY 1 LYYKNRYRYALKSGRLNAPMPENGVAENNDWVFMGYTQE 40
DB 2 LYYKNRYRYALKSGSVNAPMPENGTENNNDWITMGSTQE 41
XX
XX RESULT 7
XX AAR07304
XX AAR07304 standard; protein; 1541 AA.
XX
XX AAR07304;
AC
XX 31-JAN-1991 (first entry)
DT
XX
XX Iga1 protease.
DE
XX
XX Iga1; vaccine; meningitis; gonorrhoea; allergies.
KM
XX
XX Haemophilus influenzae.
OS
XX
XX WO9011367-A.
PN
XX
XX 04-OCT-1990.
PD
XX
XX 16-MAR-1990; 90MO-DK00073.
PF
XX
XX 17-MAR-1989; 89DK-0001308.
PR
XX
XX (KILI/) KILIAN M.
PA
XX
XX Kilian M, Poulsen K;
PI
XX
XX WPI: 1990-320267/42.
DR
XX
XX N-PSDB: AAG06164.
DR
XX
XX Immunoglobulin A1 protease prodn. - by cloning from
PT microorganisms for immunisation against immunoglobulin A1
PT protease producing bacteria
XX
XX
XX Disclosure; fig 3; 44pp; English.
PS
XX
XX This immunoglobulin (IgA) protease is produced by recombinant DNA
CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,
CC gonorrhoea or allergic diseases. It specifically cleaves the heavy

```

```

CC chain of human Iga1 in the hinge region.
XX
XX Sequence 1541 AA;
SQ
XX
XX Query Match 41.1%; Score 92; DB 11; Length 1541;
XX Best Local Similarity 37.5%; Pred. No. 0.0016;
XX Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
OY 1 LYYKNRYRYALKSGRLNAPMPENGVAENNDWVFMGYTQE 40
DB 595 LNEINTTYALKRGKASTRESEFNSENNDWIMKTSID 634
XX
XX RESULT 8
XX AAY34542
XX AAY34542 standard; Protein; 834 AA.
XX
XX AAY34542;
AC
XX
XX 25-AUG-1999 (first entry)
DT
XX
XX Porphyromonas gingivalis protein PG71.
DE
XX
XX Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KM Porphyromonas gingivalis;
KW vaccine; antigenic.
XX
XX Porphyromonas gingivalis.
OS
XX
XX WO9929870-A1.
PN
XX
XX 17-JUN-1999.
PD
XX
XX 10-DEC-1998; 96MO-AU01023.
PF
XX
XX 04-AUG-1998; 98AU-0005028.
PR 10-DEC-1997; 97AU-0000839.
PR 31-DEC-1997; 97AU-0001182.
PR 30-JAN-1998; 98AU-0001546.
PR 10-MAR-1998; 98AU-0002264.
PR 09-APR-1998; 98AU-0002911.
PR 23-APR-1998; 98AU-0003128.
PR 05-MAY-1998; 98AU-0003338.
PR 22-MAY-1998; 98AU-0003654.
PR 29-JUL-1998; 98AU-0004917.
XX
XX (CSLC-) CSL LTD.
PA
XX
XX Agius CT, Barr IG, Hocking DM, Margel's MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
PI
XX
XX WPI: 1999-385613/32.
DR
XX
XX N-PSDB: AAX91760.
DR
XX
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
PT
XX
XX Claim 1; Page 535-536; 588pp; English.
PS
XX
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (Pg) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX Sequence 834 AA;
SQ
XX
XX Query Match 28.1%; Score 63; DB 20; Length 834;
XX Best Local Similarity 36.5%; Pred. No. 7.2;
XX Matches 19; Conservative 6; Mismatches 11; Indels 16; Gaps 4;

```



```
XX 14-FEB-2002 (first entry)
DT
XX
DE E. coli cellular proliferation protein #15.
XX
KW Antisense: prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Escherichia coli.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS52293.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 10027; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 452 AA:
Query Match 27.2%; Score 61; DB 22; Length 452;
Best Local Similarity 42.3%; Pred. No. 6.6;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;
```

```
XX 14-FEB-2002 (first entry)
DT
XX
DE Salmonella typhi cellular proliferation protein #372.
XX
KW Antisense: prokaryotic cellular proliferation protein;
KM antibiotic; antibacterial; drug design.
XX
OS Salmonella typhi.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS56340.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
PS Example 3; Seq ID No 14074; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 452 AA:
Query Match 27.2%; Score 61; DB 22; Length 452;
Best Local Similarity 42.3%; Pred. No. 6.6;
Matches 11; Conservative 6; Mismatches 5; Indels 4; Gaps 1;
```


XX		21-SEP-2001	(first entry)	
DT				
XX		Escherichia coli protein sequence SEQ ID NO:450.		
DE				
XX		Escherichia coli; identification; proliferation; microorganism;		
KW		antimicrobial; antibacterial; antibiotic; gene therapy; diagnosis;		
KW		bacterial growth inhibition.		
XX				
OS		Escherichia coli.		
PN		MO200148209-A2.		
PD		05-JUL-2001.		
XX				
PF		19-DEC-2000; 2000WO-US34419.		
PR		23-DEC-1999; 99US-0173005.		
PA		(ELIT-) ELITRA PHARM INC.		
PI		Forsyth RA, Ohlsen KL, Zyskind JW;		
DR		WPI: 2001-457376/49.		
DR		N-PSDB; AAH81458.		
PT		Novel nucleic acids encoding proteins required for Escherichia coli		
XX		proliferation, useful for screening for antimicrobial agents -		
XX		Claim 19; Page 569; 596pp; English.		
XX		The present invention describes a purified or isolated nucleic acid		
CC		sequence (I) consisting essentially of one of the 93 nucleotide sequences		
CC		given in AAH81202 to AAH81294, where expression of the nucleic acid in a		
CC		microorganism is capable of inhibiting proliferation of a microorganism.		
CC		(I) have antibacterial and antibiotic activities, and can be used in		
CC		gene therapy. Expression of (I) in a microorganism inhibits proliferation		
CC		of the microorganism, and the manufactured antibiotic is useful for		
CC		reducing the activity or level of a gene product required for		
CC		proliferation of a microorganism in a subject, specifically humans. The		
CC		nucleic acids that inhibit bacterial growth or proliferation can be used		
CC		as antisense therapeutics for killing bacteria. In addition to		
CC		therapeutic applications, the nucleic acid sequences complementary to		
CC		sequences required for proliferation can be used as diagnostic tools.		
CC		For example, nucleic acid probes complementary to proliferation-required		
CC		sequences that are specific for particular species of microorganisms can		
CC		be used as probes to identify particular microorganism species in		
CC		clinical specimens. AAH81295 to AAH81487 encode the Escherichia coli		
CC		proteins given in AAG98439 to AAG98491, and AAH81488 to AAH81491		
CC		represent oligonucleotides, which are used in the exemplification of the		
CC		present invention.		
SQ		Sequence 452 AA;		
OY		Query Match 27.2%; Score 61; DB 22; Length 452;		
		Best Local Similarity 42.3%; Pred. No. 6.6;		
Matches	11; Conservative 7; Mismatches 4; Indels 4; Gaps 1			
DB		12 KSGRLNAMPENG-V---AENNDDV 33		
		: : : : : : : : : :		
ID		AAWS0891		
ID		AAWS0891 standard; Protein: 3084 AA.		
AC		AAWS0891;		
XX				
DT		07-DEC-1998 (first entry)		
DE		Mouse laminin A chain.		

KW	Laminin mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
KW	Down's syndrome; hereditary cerebral hemorrhage; inflammation;
KW	malignancy; Familial Mediterranean Fever; multiple myeloma;
KW	Type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
KW	Gertschman-Straussler syndrome; kuru; scrapie; haemodialysis;
KW	carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
KM	Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
KW	Therapy.
XX	
OS	Mus sp.
XX	
FH	Key Location/Qualifiers
FT	Domain 2746..2922
FT	/note= "fourth globular domain repeat (Claim 13)"
FT	Region 2690..2700
FT	/note= "beta-amyloid protein binding region
FT	(claim 12)"
XX	
PN	W09815179-A1.
PD	
PP	16-APR-1998.
PE	08-OCT-1997; 97WO-US18145.
PX	
PX	08-OCT-1996; 96US-0027981.
PR	
PX	08-OCT-1996; 96US-0027981.
PA	(UNIW) UNIV WASHINGTON.
PX	
PI	Castillo G, Snow AD;
PX	
DR	WPt: 1998-240534/21.
PX	
PT	Use of laminin and fragments - for developing products for use in
PT	the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
PT	disease or CJD
PX	
PS	Claim 15; Page 74-79; 132pp; English.
XX	
CC	This is the amino acid sequence of the mouse laminin A chain. The
CC	primary object of the invention is to use laminin, laminin-derived
CC	protein fragments and/or laminin-derived polypeptides as potent
CC	inhibitors of amyloid formation, deposition, accumulation and/or
CC	persistence in Alzheimer's disease and other amyloidoses. The
CC	Laminin products (see AAW50888-98) may include mouse or human laminin
CC	A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
CC	laminin G1 chain, the globular repeats of the laminin A1 chain and
CC	the beta-amyloid binding domain of the laminin A chain. A claimed
CC	method for treating an amyloid disease comprises administering a
CC	polypeptide having a conformational similarity to a fragment of a
CC	laminin protein. A method for diagnosing an amyloid disease
CC	involves determining levels of laminin in a sample. Production
CC	of laminin or its fourth globular repeat in vivo provides a method
CC	for in vivo inhibition of beta-amyloid amyloidosis. The products
CC	and methods can be used for the diagnosis, prognosis, monitoring
CC	and treatment of amyloidoses such as Alzheimer's disease, Down's
CC	syndrome and hereditary cerebral haemorrhage with amyloidosis of
CC	the Dutch type (where the specific amyloid is the beta-amyloid
CC	protein), the amyloidosis associated with chronic inflammation,
CC	various forms of malignancy and Familial Mediterranean Fever (AA
CC	amyloid or inflammation-associated amyloidosis), the amyloidosis
CC	associated with multiple myeloma and other B-cell abnormalities
CC	(AL amyloid), the amyloidosis associated with type II diabetes
CC	(amylin or islet amyloid), the amyloidosis associated with prion
CC	diseases including Creutzfeldt-Jacob disease, Gertschman-Straussler
CC	syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
CC	associated with long-term haemodialysis and carpal tunnel syndrome
CC	(beta 2-microglobulin amyloid), the amyloidosis associated with
CC	senile cardiac amyloid and Familial Amyloidotic Polynuropathy
CC	(presenium or transylretein amyloid), and the amyloidosis
CC	associated with endocrine tumours such as medullary carcinoma of
CC	the thyroid (variant of procalcitonin).
XX	
SQ	Sequence 3084 AA:

Query Match 25.9%; Score 58; DB 19; Length 3084;
 Best Local Similarity 30.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 14; Gaps 2;

OY 5 NRYRYALKSGR-----LNAPMPENGVAENND-----WVFMGYTOE 40
 DB 1260 NYEPQVLKGGRRKHVITYMDAPAPENGVRQDYEVQMKKEEFKYNFSVSE 1309

RESULT 15
 AAEL1215

ID AAEL1215 standard; Protein; 3084 AA.

XX AAEL1215;

DT 03-JAN-2002 (first entry)

DE Mouse laminin-1 alpha chain protein.

XX Mouse; laminin-1; alpha chain; matn; cytosolic; ophthalmological;
 KW vulneryary; angiogenesis-mediated disease; psoriasis; obesity; vasotropic;
 KW haemostatic; diabetic retinopathy; angiofibroma; anorectic; gene therapy;
 KW dermatological; antinflammatory; contraception; cancer; benign tumour;
 KW rheumatoid arthritis; endothelial cell proliferation; atherosclerosis;
 KW fibrosis; wound granulation; intestinal adhesion; Crohn's disease;
 KW scleroderma; Helicobacter pylori ulcer; contraception.

OS Mus musculus.

XX Key location/Qualifiers

FT Domain

FT 2132..2338 /label=G1_domain

FT /note="Globular domain 1"

FT 2336..2517 /label=G2_domain

FT /note="Globular domain 2"

FT 2518..2745 /label=G3_domain

FT /note="Globular domain 3"

FT 2746..2879 /label=G4_domain

FT /note="Globular domain 4"

FT 2880..3084 /label=G5_domain

FT /note="Globular domain 5"

FT Domain

FT WO200173033-A2.

XX 04-OCT-2001.

XX 28-MAR-2001; 2001WO-US09921.

XX 29-MAR-2000; 2000US-192875P.

XX (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.

XX Kalluri R;

XX WPI: 2001-639233/73.

XX N-PSDB; AAD20143.

XX New proteins comprising the globular domains of alpha chain of laminin

XX having anti-angiogenic properties for treating angiogenesis-mediated

XX diseases such as cancer, psoriasis, ulcer, rheumatoid arthritis and

XX obesity

XX Claim 4; Fig 1; 78pp; English.

XX The invention relates to Matn an isolated protein of globular 1 domain

XX of alpha-chain of laminin having antiangiogenic activity. Matn and

XX laminin proteins are useful for treating a disorder especially tumour

XX growth involving inhibiting angiogenesis by inhibiting one or more of

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

CC endothelial cell proliferation, endothelial cell migration or endothelial
 CC cell tube formation in a tissue and also for treating a disorder by
 CC promoting or inducing endothelial cell apoptosis in a tissue. Matn and
 CC laminin proteins are used to inhibit angiogenic activity characteristic
 CC of a disease such as angiogenesis-dependent cancers, benign tumours,
 CC rheumatoid arthritis, diabetic retinopathy, fibrosis, psoriasis, ocular
 CC angiogenesis diseases, Osler-Webber Syndrome, myocardial angiogenesis,
 CC plaque neovascularisation, telangiectasia, haemophilic joints,
 CC angiofibroma, wound granulation, intestinal adhesions, atherosclerosis,
 CC scleroderma, hypertrophic scars, cat scratch disease, Helicobacter
 CC pylori ulcers, dialysis graft vascular access stenosis, contraception,
 CC obesity and also Crohn's disease. Laminin DNA is useful in gene
 CC therapy and also for designing probes to isolate the anti-angiogenic
 CC proteins. The present sequence is mouse laminin-1 alpha chain protein.

SO Sequence 3084 AA;

Query Match 25.9%; Score 58; DB 22; Length 3084;
 Best Local Similarity 30.0%; Pred. No. 1.6e+02;
 Matches 15; Conservative 6; Mismatches 15; Indels 14; Gaps 2;

OY 5 NRYRYALKSGR-----LNAPMPENGVAENND-----WVFMGYTOE 40
 DB 1260 NYEPQVLKGGRRKHVITYMDAPAPENGVRQDYEVQMKKEEFKYNFSVSE 1309

Search completed: December 10, 2002, 10:54:05
 Job time : 29.2 secs

```

RESULT 2
US-09-901-252-15
: Sequence 15, Application US/09901252
: Patent No. US20020026658A1
: GENERAL INFORMATION:
: APPLICANT: Chapple, Clint
: TITLE OF INVENTION: Genes Encoding Sinapoylglucose:malate Sinapoyltransferase and
: TITLE OF INVENTION: Use
: FILE REFERENCE: N1422-004
: CURRENT APPLICATION NUMBER: US/09/901,252
: CURRENT FILING DATE: 2001-07-09

```


	Matches	11; Conservative	6; Mismatches	5; Indels	4; Gaps	1;
Qy	12	KSGGRINAPMPENCV----	AENNDAV	33		
		1: 1: : : 1: :				
Db	200	KAKGEITGTPENGAIINMADNNML	225			

RESULT 6

```

US-09-938-275-4
? Sequence 4, Application US/09938275
? Patent No. US2002011309A1
? GENERAL INFORMATION:
? APPLICANT: Gerardo Castillo
? APPLICANT: Alan Snow
? TITLE OF INVENTION: Therapeutic and Diagnostic Applications
? TITLE OF INVENTION: of Lamlnin and Lamlnin-Derived Protein Fragments
? FILE REFERENCE: PROTO.P03
? CURRENT APPLICATION NUMBER: US/09/938,275
? CURRENT FILING DATE: 2001-08-16
? NUMBER OF SEQ ID NOS: 11
? SOFTWARE: FastSeq for Windows Version 4.0
? SEQ ID NO 4
? LENGTH: 3084
? TYPE: PRT
? ORGANISM: Mus Musculus
? PUBLICATION INFORMATION:
? DATABASE ACCESSION NUMBER: Swissprot P19137
? DATABASE ENTRY DATE: 1990-11-01
? US-09-938-275-4

```

Query Match	25.9%	Score	58	DB	10	Length	3084
Best Local Similarity	30.0%	Pred. NO.	12				
Matches	15	Conservative	6	Mismatches	15	Indels	14
						Gaps	2

QY 5 NRYRYALKSGGR-----LNAPMPENGVAENND-----WVFMGYTQE 40
 || | ||| ::|| |||| ::| :
 Db 1260 WNEPQVLIKGRARKHVIYMDAPAPENCYRQDYEVQMKKEEPWKYFNSVSE 1309

RESULT 7

```

US-09-938-275-5
? Sequence 5, Application US/09938275
? Patent No. US20020111309A1
?
? GENERAL INFORMATION:
?
? APPLICANT: Gerardo Castillo
?
? APPLICANT: Alan Snow
?
? TITLE OF INVENTION: Therapeutic and Diagnostic Applications
?
? TITLE OF INVENTION: of Laminin and Laminin-Derived Protein
?
? FILE REFERENCE: PROTEO_P03
?
? CURRENT APPLICATION NUMBER: US/09/938,275
?
? CURRENT FILING DATE: 2001-08-16
?
? NUMBER OF SEQ ID NOS: 11
?
? SOFTWARE: FastSeq For Windows Version 4.0
?
? SEQ ID NO 5
?
? LENGTH: 3075
?
? TYPE: PRT
?
? ORGANISM: Homo Sapiens
?
? PUBLICATION INFORMATION:
?
? DATABASE ACCESSION NUMBER: Swissprot P25391
?
? DATABASE ENTRY DATE: 1992-05-01
?
? US-09-938-275-5

```

Query Match	25.0%	Score	56	DB	10	Length	3075
Best Local Similarity	30.0%	Pred.	No. 23				
Matches	15	Conservative	5	Mismatches	16	Indels	14
						Gaps	2

```
QY      5  NYRYALKSGGR-----LNAPMEENGVAE-----NNDWVFMGYTQE 40
      1:  |  ||  ::||  ||||  :  |  |  |
Db 1253 NEPOVLIKGRIRKQVIYMDAPAPENGVRQEQEVAMRENFWKFNYSVE 1302
```

RESULT 8
US-09-765-272-154

Sequence 154, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
ADDITIVE: 0.1

; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Hwang Gwang, College, Tae

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Bethesda

CITY: ROCKVILLE
STATE: Maryland
COUNTRY: USA

COUNTRY: USA
ZIP: 20850
COUNTY: DISTRICT OF COLUMBIA

```
;
;      COMPUTER READABLE FORM:
;
;      MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;      Generated by: IBM PC DOS 6.22
```

COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2

```

; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001

```

; .
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:

```

APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anderson

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PEE

TELECOMMUNICATION INFORMATION
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8543

TELEFAX: (301) 309-8525
; INFORMATION FOR SEQ ID NO: 154:
; CDS: 1..1000

SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids

```

;
; TYPE: amino acid
STRANDEDNESS: single
MODIFIER: 1

```

MOLECULE TYPE: protein
SOURCE: DECEMBER

SEQUENCE DESCRIPTION: SEQ ID NO: 154:
US-09-765-272-154

Query Match	24.38;	Score 54.5;	DB 10;	Length 569;
-------------	--------	-------------	--------	-------------

Best Local Similarity 27.1%; Pred. NO. 5.4;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1.

0Y 3 YKNRYALKSGGRINAP-----MPENGVAENNDWVEMGYTQ 39

Db 129 FENGHYYYLKS GGYMAANEWIMDKESWFLKFDGMAEK EWVYDSHSQ 176

RESULT 9

US-09-/65-212-/4
; Sequence 74, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:

```
;; APPLICANT: Choi et. al.  
;; TITLE OF INVENTION: Str
```

NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue

```

; CITY: Rockville
; STATE: Maryland

```

```

;
; COUNTRY: USA
; ZIP: 20850

```

```

;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

```

```
;
;      COMPUTER: HP Vectra 486/33
;      OPERATING SYSTEM: MSDOS version 6.2
```

```

;      SOFTWARE: ASCII Text
;
;      CURRENT APPLICATION DATA:

```

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 74:
SEQUENCE CHARACTERISTICS:
LENGTH: 591 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 74:
US-09-765-272-74

Query Match 24.3% Score 54.5; DB 10; Length 591;
Best Local Similarity 27.1% Pred. No. 5.6;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

QY 3 YKNRYVALKSGRLNAP-----MPENGVAENNDVFMGYTQ 39
DB 190 FENGHYLYLKSGGYMANEMIMKESFYLPFGKMAEKRYVDHSQ 237

RESULT 10
US-09-815-242-5118
Sequence 5118, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA 011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5118
LENGTH: 1317
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5118

Query Match 22.1% Score 49.5; DB 10; Length 1317;
Best Local Similarity 33.3% Pred. No. 71;

Matches 12; Conservative 4; Mismatches 9; Indels 11; Gaps 1;
QY 1 LYKNRYVALKSGRLNAPMPENGVAENNDVFMG 36
DB 1122 LYNNRYRY-----LPEAGRYASODPLGLG 1146

RESULT 11
US-09-879-957-24
Sequence 24, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.
HOFFMAN, No. US20020034755A1h
KAY, Brian K.
FOWLKES, Dana M.
MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/879,957
FILING DATE: 13-Jun-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/630,915
FILING DATE: 03-Apr-1996
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 404 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: unknown
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-879-957-24

Query Match 21.9% Score 49; DB 10; Length 404;
Best Local Similarity 52.9% Pred. No. 22;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 24 NGVAENNDVFMGYTQ 40
DB 66 NKIAENNDLIMDYHOK 82

RESULT 12
US-09-879-957-22
Sequence 22, Application US/09879957
Patent No. US20020034755A1
GENERAL INFORMATION:
APPLICANT: SPARKS, Andrew B.

```

;
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FOMKES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
;
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIORITY DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
;
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 434 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-09-879-957-22

Query Match 21.9%; Score 49; DB 10; Length 434;
Best Local Similarity 52.9%; Pred. No. 23;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 24 NGVAENNDWFMGYTOE 40
| : |||| : || : |
Db 101 NRIKENNDLMDYHOK 117

RESULT 13
US-09-998-598-2603
; Sequence 2603, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madeline Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2603
; LENGTH: 523
; TYPE: PRT
```

```

; ORGANISM: Homo sapiens
US-09-998-598-2603

Query Match 21.7%; Score 48.5; DB 10; Length 523;
Best Local Similarity 55.6%; Pred. No. 34;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

OY 5 NRYRYALSGGRINAPMP 22
| : |||| : || : |
Db 407 NTRFFA-QSGGRINPLP 423

RESULT 14
US-09-966-561-2
; Sequence 2, Application US/09966561
; Patent No. US20020090696A1
; GENERAL INFORMATION:
; APPLICANT: University of Southern California
; APPLICANT: Miller, Carol A.
; APPLICANT: Dong, Zhao Hui
; APPLICANT: Zhang, Yan
; TITLE OF INVENTION: APOPTOSIS INHIBITION
; FILE REFERENCE: 13761-724
; CURRENT APPLICATION NUMBER: US/09/966,561
; CURRENT FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/419,694
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-10-14
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 711
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-966-561-2

Query Match 21.7%; Score 48.5; DB 10; Length 711;
Best Local Similarity 29.4%; Pred. No. 48;
Matches 15; Conservative 8; Mismatches 11; Indels 17; Gaps 3;

OY 6 YRYIALKSGGRINAPM-----PEN--GVAENNDW-----FMGYTO 39
| : |||| : || : |
Db 526 YEAVNMRTGARGVFPAYVAIEVTKPEPHMALAKNSDWDVQFRVKFLGSVO 576

RESULT 15
US-09-873-737A-6
; Sequence 6, Application US/09873737A
; Patent No. US20020076797A1
; GENERAL INFORMATION:
; APPLICANT: Duke University
; APPLICANT: Lin, Haifan
; TITLE OF INVENTION: PURIFIED AND ISOLATED piwi FAMILY GENES AND GENE
; FILE REFERENCE: PRODUCTS AND THERAPEUTIC AND SCREENING METHODS USING SAME
; CURRENT APPLICATION NUMBER: US/09/873,737A
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: PCT/US99/28764
; PRIOR FILING DATE: 1999-12-03
; PRIOR APPLICATION NUMBER: 60/110,901
; PRIOR FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 861
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (76)
; OTHER INFORMATION: Xaa-Leu or Ile
; NAME/KEY: misc_feature
; LOCATION: (303)
; OTHER INFORMATION: Xaa-Leu or Ile
```

; NAME/KEY: misc_feature
; LOCATION: (735)
; OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-6

Query Match 21.7%; Score 48.5; DB 10; Length 861;
Best Local Similarity 55.6%; Pred. No. 60;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
QY 5 NYRYALKSGGRLNAPMP 22
| | : | : | | | | | | : | : |
Db 745 NTRFFA-QSGGRLONPLP 761

Search completed: December 10, 2002, 10:58:09
Job time : 6.2 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:54:49 ; Search time 9.8 Seconds
(without alignments)
268.490 Million cell updates/sec

Title: US-09-142-970-4_COPY_1_40
Perfect score: 224
Sequence: 1 LYKKNRYALKSGRLNAPMPENGVAENNDWFMGYTOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 232864 seqs, 65780094 residues

Total number of hits satisfying chosen parameters: 232864

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/PCOT_NEW_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	25.0	2606	5	US-09-724-676-72290 Sequence 72290, A
2	56	25.0	2606	5	US-09-724-676A-72290 Sequence 72290, A
3	52.5	23.4	62	6	US-10-209-582-1019 Sequence 1019, Ap
4	51	22.8	82	6	US-10-209-582-657 Sequence 657, App
5	50.5	22.5	378	6	US-10-216-209-87 Sequence 87, Appl
6	50.5	22.5	439	5	US-09-134-000C-6557 Sequence 6557, Ap
7	50	22.3	349	1	PCT-US02-32727-7883 Sequence 7883, Ap
8	50	22.3	349	6	US-10-057-498-7883 Sequence 7883, Ap
9	49	21.9	329	5	US-09-724-676-75420 Sequence 75420, A
10	49	21.9	329	5	US-09-724-676-75421 Sequence 75421, A
11	49	21.9	329	5	US-09-724-676A-75420 Sequence 75420, A
12	49	21.9	329	5	US-09-724-676A-75421 Sequence 75421, A
13	49	21.9	365	5	US-09-724-676-75392 Sequence 75392, A
14	49	21.9	365	5	US-09-724-676-75393 Sequence 75393, A
15	49	21.9	365	5	US-09-724-676A-75392 Sequence 75392, A
16	49	21.9	365	5	US-09-724-676A-75393 Sequence 75393, A
17	49	21.9	385	5	US-09-724-676-75418 Sequence 75418, A
18	49	21.9	385	5	US-09-724-676-75419 Sequence 75419, A
19	49	21.9	385	5	US-09-724-676A-75418 Sequence 75418, A
20	49	21.9	385	5	US-09-724-676A-75419 Sequence 75419, A
21	49	21.9	414	5	US-09-724-676-75400 Sequence 75400, A
22	49	21.9	414	5	US-09-724-676-75401 Sequence 75401, A
23	49	21.9	414	5	US-09-724-676-75402 Sequence 75402, A
24	49	21.9	414	5	US-09-724-676-75403 Sequence 75403, A
25	49	21.9	414	5	US-09-724-676A-75400 Sequence 75400, A
26	49	21.9	414	5	US-09-724-676A-75401 Sequence 75401, A

27	49	21.9	414	5	US-09-724-676A-75402 Sequence 75402, A
28	49	21.9	414	5	US-09-724-676A-75403 Sequence 75403, A
29	49	21.9	415	5	US-09-724-676-75411 Sequence 75411, A
30	49	21.9	415	5	US-09-724-676-75412 Sequence 75412, A
31	49	21.9	415	5	US-09-724-676A-75411 Sequence 75411, A
32	49	21.9	415	5	US-09-724-676A-75412 Sequence 75412, A
33	49	21.9	439	5	US-09-724-676-75404 Sequence 75404, A
34	49	21.9	439	5	US-09-724-676-75405 Sequence 75405, A
35	49	21.9	439	5	US-09-724-676-75413 Sequence 75413, A
36	49	21.9	439	5	US-09-724-676-75414 Sequence 75414, A
37	49	21.9	439	5	US-09-724-676-75416 Sequence 75416, A
38	49	21.9	439	5	US-09-724-676-75417 Sequence 75417, A
39	49	21.9	439	5	US-09-724-676A-75404 Sequence 75404, A
40	49	21.9	439	5	US-09-724-676A-75405 Sequence 75405, A
41	49	21.9	439	5	US-09-724-676A-75413 Sequence 75413, A
42	49	21.9	439	5	US-09-724-676A-75414 Sequence 75414, A
43	49	21.9	439	5	US-09-724-676A-75416 Sequence 75416, A
44	49	21.9	439	5	US-09-724-676A-75417 Sequence 75417, A
45	49	21.9	458	5	US-09-724-676-75386 Sequence 75386, A

ALIGNMENTS

```
RESULT 1
US-09-724-676-72290
; Sequence 72290, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724, 676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 72290
; LENGTH: 2606
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-72290

Query Match          25.0%; Score 56; DB 5; Length 2606;
Best Local Similarity 30.0%; Pred. No. 26;
Matches 15; Conservative 5; Mismatches 16; Indels 14; Gaps 2;

OY      5 NYRYALKSGR-----LNAPMPENGVAE-----NNDWFMGYTOE 40
       1:      |||      ::|| |||||      | :
Db      1253 NPEPOVLKIGRIKQVIYMDAPAPENGVAEENNDWFMGYTSYSE 1302
       1:      |||      ::|| |||||      | :

RESULT 2
US-09-724-676A-72290
; Sequence 72290, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Comugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Comugen
; CURRENT APPLICATION NUMBER: US/09/724, 676A
; CURRENT FILING DATE: 2000-11-28
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 72290
; LENGTH: 2606
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-72290

Query Match          25.0%; Score 56; DB 5; Length 2606;
Best Local Similarity 30.0%; Pred. No. 26;
Matches 15; Conservative 5; Mismatches 16; Indels 14; Gaps 2;

OY      5 NYRYALKSGR-----LNAPMPENGVAE-----NNDWFMGYTOE 40
       1:      |||      ::|| |||||      | :

```

Db 1253 NEEPOYLINGGIRIKOVIYMDAPAPENGVRQEOEVAMRENFWKENSASE 1302

RESULT 3
US-10-209-582-1019

; Sequence 1019, Application US/10209582

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PM009CIN

; CURRENT APPLICATION NUMBER: US/10/209,582

; PRIOR FILING DATE: 2002-08-01

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; NUMBER OF SEQ ID NOS: 1030

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1019

; LENGTH: 62

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (47)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (62)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; US-10-209-582-1019

Query Match

Best Local Similarity 23.4%; Score 52.5; DB 6; Length 62;

Matches 12; Conservative 8; Mismatches 13; Indels 1; Gaps 1;

Qy 7 RYALKSGRLNAPMPENGVAENNDWVFMGYTQ 40

Db 14 RGTFLKRGHTTASPEESGVGME-VFLGYSFE 46

RESULT 4
US-10-209-582-657

; Sequence 657, Application US/10209582

; GENERAL INFORMATION:

; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PM009CIN

; CURRENT APPLICATION NUMBER: US/10/209,582

; PRIOR FILING DATE: 2002-08-01

; PRIOR APPLICATION NUMBER: 09/758,461

; PRIOR FILING DATE: 2001-01-11

; PRIOR APPLICATION NUMBER: 60/179,065

; PRIOR FILING DATE: 2000-01-31

; PRIOR APPLICATION NUMBER: 60/180,628

; PRIOR FILING DATE: 2000-02-04

; NUMBER OF SEQ ID NOS: 1030

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 657

; LENGTH: 82

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (37)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; US-10-209-582-657

Query Match

Best Local Similarity 22.8%; Score 51; DB 6; Length 82;

Matches 13; Conservative 4; Mismatches 10; Indels 6; Gaps 1;

Qy 7 RYALKSGRLNAPMPENGVAENNDWVFMGYTQ 39

Db 42 RPRVLSTNGKLNKAKRKRYV-----VEKALTLQ 68

RESULT 5
US-10-216-209-87

; Sequence 87, Application US/10216209

; GENERAL INFORMATION:

; APPLICANT: Lam, Joseph S.

; APPLICANT: Burrows, Lori

; APPLICANT: Charter, Deborah

; APPLICANT: De Kievit, Teresa De

; TITLE OF INVENTION: Novel Proteins Involved in the Synthesis and Assembly

; TITLE OF INVENTION: of O-Antigen in Pseudomonas Aeruginosa

; FILE REFERENCE: 6580-167

; CURRENT APPLICATION NUMBER: US/10/216,209

; PRIOR FILING DATE: 2002-08-12

; PRIOR APPLICATION NUMBER: US/09/352,994

; PRIOR FILING DATE: 2001-05-29

; PRIOR APPLICATION NUMBER: US 08/846,762

; PRIOR FILING DATE: 1997-04-30

; NUMBER OF SEQ ID NOS: 100

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 87

; LENGTH: 378

; TYPE: PRT

; ORGANISM: Salmonella enterica

; US-10-216-209-87

Query Match

Best Local Similarity 22.5%; Score 50.5; DB 6; Length 378;

Matches 9; Conservative 9; Mismatches 6; Indels 3; Gaps 1;

Qy 2 YKKNRYALKSGR---LNAPMPENG 25

Db 108 FYQNIKWIEAGLRTWMNSPPEEG 134

RESULT 6
US-09-134-000C-6557

; Sequence 6557, Application US/09134000C

; GENERAL INFORMATION:

; APPLICANT: Lynn Doucette-Stamm et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO

; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 032796-032

; CURRENT APPLICATION NUMBER: US/09/134,000C

; PRIOR FILING DATE: 1998-08-13

; PRIOR APPLICATION NUMBER: US 60/055,778

; PRIOR FILING DATE: 1997-08-15

; NUMBER OF SEQ ID NOS: 6812

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 6557

; LENGTH: 439

; TYPE: PRT

; ORGANISM: Enterococcus faecalis

; FEATURE:

; NAME/KEY: MISC_FEATURE

; LOCATION: (439)..(439)

; OTHER INFORMATION: Amino acid 439 is Xaa wherein Xaa = any amino acid.

; US-09-134-000C-6557

Query Match

Best Local Similarity 22.5%; Score 50.5; DB 5; Length 439;

Matches 13; Conservative 5; Mismatches 11; Indels 9; Gaps 1;

Qy 8 YALKSGRLNAPMPENGVAENNDWVFMG 36

Db 367 YTYKSGDTLKNIAQYGVANLRSMNGISGLIFVG 404

```
RESULT 7
PCT-US02-32727-7883
; Sequence 7883, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Mitchell, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30992
; SEQ ID NO 7883
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Propionibacterium acnes
PCT-US02-32727-7883

Query Match          22.3%; Score 50; DB 1; Length 349;
Best Local Similarity 38.5%; Pred. No. 21;
Matches 10; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY      8 YYALKSGRLNAPMPENGVAENNDWY 33
      : | | | | : | : | | | |
      : | | | | : | : | | | |

DB      313 FSADKYGKNGKRAVEWISRRNDWY 338

RESULT 8
US-10-057-498-7883
; Sequence 7883, Application US/10057498
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skelky, Yasir
; APPLICANT: Persing, David
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acnes
; FILE REFERENCE: 210121.514
; CURRENT APPLICATION NUMBER: US/10/057,498
; CURRENT FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 29212
; SEQ ID NO 7883
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Propionibacterium acnes
US-10-057-498-7883

Query Match          22.3%; Score 50; DB 6; Length 349;
Best Local Similarity 38.5%; Pred. No. 21;
Matches 10; Conservative 4; Mismatches 12; Indels 0; Gaps 0;

QY      8 YYALKSGRLNAPMPENGVAENNDWY 33
      : | | | | : | : | | | |
      : | | | | : | : | | | |

DB      313 FSADKYGKNGKRAVEWISRRNDWY 338

RESULT 9
US-09-724-676-75420
; Sequence 75420, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
```

```
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75420
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75420

Query Match          21.9%; Score 49; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      24 NGVAENNDWVFNGYTOE 40
      | : | | | | : | | | : |
      | : | | | | : | | | : |

DB      101 NKIAENNDLMDYHOK 117

RESULT 10
US-09-724-676-75421
; Sequence 75421, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75421
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75421

Query Match          21.9%; Score 49; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      24 NGVAENNDWVFNGYTOE 40
      | : | | | | : | | | : |
      | : | | | | : | | | : |

DB      101 NKIAENNDLMDYHOK 117

RESULT 11
US-09-724-676A-75420
; Sequence 75420, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75420
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-75420

Query Match          21.9%; Score 49; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY      24 NGVAENNDWVFNGYTOE 40
      | : | | | | : | | | : |
      | : | | | | : | | | : |

DB      101 NKIAENNDLMDYHOK 117

RESULT 12
US-09-724-676A-75421
; Sequence 75421, Application US/09724676A
```

```
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75421
; LENGTH: 329
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-75421

Query Match          21.9%; Score 49; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 27;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 24 NGVAENNDWFMGYTOE 40
   1 :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 13
US-09-724-676-75392
; Sequence 75392, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75392
; LENGTH: 365
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-75392

Query Match          21.9%; Score 49; DB 5; Length 365;
Best Local Similarity 52.9%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 24 NGVAENNDWFMGYTOE 40
   1 :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 14
US-09-724-676-75393
; Sequence 75393, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75393
; LENGTH: 365
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676-75393

Query Match          21.9%; Score 49; DB 5; Length 365;
Best Local Similarity 52.9%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 24 NGVAENNDWFMGYTOE 40
   1 :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117
```

```
RESULT 15
US-09-724-676A-75392
; Sequence 75392, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75392
; LENGTH: 365
; TYPE: PRF
; ORGANISM: Homo sapiens
US-09-724-676A-75392

Query Match          21.9%; Score 49; DB 5; Length 365;
Best Local Similarity 52.9%; Pred. No. 30;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 24 NGVAENNDWFMGYTOE 40
   1 :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117
```

Search completed: December 10, 2002, 11:10:17
Job time : 10 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:50:48 ; Search time 11 seconds
(without alignments)
349.580 Million cell updates/sec

Title: US-09-142-970-4_COPY_1_40
Perfect score: 224
Sequence: 1 LYKKNRYRYALKSGRLNAPMPENGVAENNDWFMGYTOE 40

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	99.6	1532	2 A26039	IgA-specific metal
2	189	84.4	1561	2 S61314	IgA-specific metal
3	189	84.4	1773	2 A81937	IgA-specific metal
4	185	82.6	1815	2 C81169	IgA-specific metal
5	92	41.1	1541	2 A37023	IgA-specific metal
6	92	41.1	1694	2 H64106	IgA-specific metal
7	92	41.1	1702	2 A41859	IgA-specific metal
8	84	37.5	1849	2 C41859	IgA-specific metal
9	81	36.2	1545	2 B41859	IgA-specific metal
10	62.5	27.9	532	1 C6917	IgA-specific metal
11	62	27.7	297	2 S5085	carboxypeptidase C
12	61	27.2	452	2 B90640	hypothetical prote
13	61	27.2	452	2 B85491	D-alanine-D-alanin
14	61	27.2	452	2 A10517	UDP-N-acetylmutam
15	61	27.2	452	2 P64730	UDP-N-acetylmutam
16	58	25.9	3084	1 MMMSA	lamnin alpha-1 ch
17	56	25.0	189	2 B69020	hypothetical prote
18	56	25.0	1367	2 T33819	hypothetical prote
19	56	25.0	3075	2 S14458	lamnin alpha-1 ch
20	55.5	24.8	993	2 A38437	probable homeotic
21	55	24.6	350	2 S47292	phenol 2-monooxyge
22	55	24.6	574	2 A91182	hypothetical prote
23	55	24.6	574	2 E86028	hypothetical prote
24	54.5	24.3	658	2 E95111	endo-beta-N-acetyl
25	54.5	24.3	721	2 C97980	endo-beta-N-acetyl
26	54.5	24.3	766	2 T20003	hypothetical prote
27	53.5	23.9	660	2 S70904	transferrin-biudin
28	53.5	23.9	850	2 S20462	RN12 protein - ye
29	53	23.7	1076	2 A69409	carbamoyl-phosphat

30	52.5	23.4	671	2 A38109	autolysin - Entero
31	52.5	23.4	949	2 H97322	DNA/RNA helicase,
32	52	23.2	262	2 A25140	patasporal crystal
33	52	23.2	431	2 S50633	hypothetical prote
34	52	23.2	530	2 T35201	probable transmemb
35	51.5	23.0	472	2 AD1440	hypothetical prote
36	51.5	23.0	507	2 S50949	dipeptidase (impor
37	51.5	23.0	467	2 S72269	hypothetical prote
38	51.5	23.0	467	2 A37113	ryanodine receptor
39	51.5	23.0	467	2 C64461	ryanodine receptor
40	51	22.8	821	2 C82834	DNA polymerase III
41	51	22.8	1112	2 AD0785	conserved hypothet
42	50.5	22.5	114	2 T41892	BRO-d orf2 - Bomby
43	50.5	22.5	349	2 S37048	cysteine proteins
44	50.5	22.5	447	2 A29959	DNA-directed RNA p
45	50.5	22.5	1361	2 A29959	

ALIGNMENTS

RESULT 1
A26039
Iga-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (S
N:Alterate names: Iga protease; immunoglobulin A1 proteinase
C:Species: Neisseria gonorrhoeae
A:Variety: strain MS11
C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000
C:Accession: A26039; S09386
R:Pohlner, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.
Nature 325, 458-462, 1987
A:Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae Iga prot
A:Reference number: A26039; MUID:87115823; PMID:3027577
A:Accession: A26039
A:Molecule type: DNA
A:Residues: 1-1532 <POH>
A:Cross-references: GB:X04835; NID:944868; PIDN:CAA28538.1; PID:944869
A:Note: The authors translated the codon AAG for residue 668 as Asn
R:Halter, R.; Pohlner, J.; Meyer, T.F.
EMBO J. 8, 2737-2744, 1989
A:Title: Mosaic-like organization of Iga protease genes in Neisseria gonorrhoeae gene
A:Reference number: S09386; MUID:90060036; PMID:2511009
A:Accession: S09386
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'W', 429-531, 'N', 533-615, 'V', 617-631, 'N', 6
A:Experimental source: strain MS11
C:Genetics:
A:Gene: Iga
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase; transmembrane protein
F:1-27/Domain: signal sequence #status predicted <SIC>
F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <MAT>
F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted
F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match
Best Local Similarity 99.6%; Score 223; DB 2; Length 1532;
Pred. No. 1.5e-21;
Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGRLNAPMPENGVAENNDWFMGYTOE 40
DB 584 LYKKNRYRYALKSGRLNAPMPENGVAENNDWFMGYTOE 623

RESULT 2
S61314
Iga-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis
C:Species: Neisseria meningitidis
A:Variety: HF13
C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000
C:Accession: S61314

R.Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria meningitidis*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Variety: strain MC58 serogroup B, strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <TEP>
A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AA41117.1; PID:g722
A:Experimental source: serogroup B, strain MC58
R.Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria meningitidis*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82477; NID:g732856; PIDN:CA457860.1; PID:g732857
C:Genetics: A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 84.4%; Score 189; DB 2; Length 1561;
Best Local Similarity 82.5%; Pred. No. 5.8e-17;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 LYKKNRYALKSGGRNALPMPENGVAENNDWFMGYTOE 40
|||||
584 LYKKNRYALKSGGSVNAPMPENGQTEENNNDWLMGSTOE 623

RESULT 3
A81937
Iga-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - *Neisseria meningitidis*
N:Alternate names: Iga1 protease; Iga1 protease (EC 3.4.21.7) [misnomer]; Immunoglobulin A:Variety: strain 22491 serogroup A; strain HF117; strain HF159; strain SM1027
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parthill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel, R.; Holooy, S.; Jagers, K.; Leather, S.; Mølle, S.; Mungall, K.; Quail, M.A.; Rajandream, Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* 22491.
A:Reference number: A81775; MUID:20222556; PMID:10761919
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CAB84182.1; PID:g737961
A:Experimental source: serogroup A, strain 22491
R.Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria meningitidis*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82470; NID:g732854; PIDN:CA457853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82471; NID:g732858; PIDN:CA457854.1; PID:g732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82472; NID:g732852; PIDN:CA457855.1; PID:g732853
A:Experimental source: strain SM1027
C:Genetics: A:Gene: NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 84.4%; Score 189; DB 2; Length 1773;
Best Local Similarity 82.5%; Pred. No. 6.6e-17;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 LYKKNRYALKSGGRNALPMPENGVAENNDWFMGYTOE 40
|||||
573 LYKKNRYALKSGGSVNAPMPENGQTEENNNDWLMGSTOE 612

RESULT 4
C81169
Iga-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - *Neisseria meningitidis*
C:Species: *Neisseria meningitidis*
A:Variety: strain MC58 serogroup B, strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, R.; Qin, H.; Vamathavan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <TEP>
A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AA41117.1; PID:g722
A:Experimental source: serogroup B, strain MC58
R.Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding Iga1 protease in *Neisseria meningitidis*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOM>
A:Cross-references: EMBL:X82477; NID:g732856; PIDN:CA457860.1; PID:g732857
C:Genetics: A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 82.6%; Score 185; DB 2; Length 1615;
Best Local Similarity 80.0%; Pred. No. 2.4e-16;
Matches 32; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

1 LYKKNRYALKSGGRNALPMPENGVAENNDWFMGYTOE 40
|||||
573 LYKKNRYALKSGGSVNAPMPENGQTEENNNDWLMGSTOE 612

RESULT 5
A37023
Iga-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - *Haemophilus influenzae*
N:Alternate names: immunoglobulin A1 protease type 1
C:Species: *Haemophilus influenzae*
C:Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C:Accession: A37023
R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Kilian, M. Infect. Immun. 57, 3097-3105, 1989
A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of *Haemophilus influenzae*
A:Reference number: A37023; MUID:89379374; PMID:2506130
A:Accession: A37023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1541 <POU>
A:Cross-references: GB:X64357; NID:g43560; PIDN:CA445708.1; PID:g43561
A:Experimental source: serotype b
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloprotease

Query Match 41.1%; Score 92; DB 2; Length 1541;
Best Local Similarity 37.5%; Pred. No. 0.00068;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

1 LYKKNRYALKSGGRNALPMPENGVAENNDWFMGYTOE 40
|||||
595 LNLENTTYALKRGASTRELEFRNKGSENNMLYMKRTSD 634

RESULT 6

H64106

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain N:Alternate names: immunoglobulin A1 proteinase type 1

C:Species: Haemophilus influenzae

A:Variety: strain Rd KW20

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000

A:Accession: H64106; A41500

R:Flischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

J. D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64106

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1694 <TIGR>

A:Cross-references: GB:U32779; GB:L42023; NID:g1574009; PIDN:AA22651.1; PID:g1574019; T

A:Experimental source: strain Rd KW20

R:Grundy, F.J.; Plant, A.G.; Wright, A.

Infect. Immun. 58, 320-331, 1990

A:Title: Localization of the cleavage site specificity determinant of Haemophilus influe

A:Reference number: A41500; MUID:90129281; PMID:2105270

A:Accession: A41500

A:Molecule type: DNA

A:Residues: 1-377 <GRU>

A:Cross-references: GB:X59800

A:Experimental source: strain Rd KW20

A:Note: the authors translated the codon TGG for residue 319 as Thr

C:Function:

A:Description: this proteinase is classified as type 1 because it cleaves at a proline-s

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 41.1%; Score 92; DB 2; Length 1694;

Best Local Similarity 37.5%; Pred. No. 0.0075;

Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYKRYRYALKSGRLNMPENGVAENNDVWFGTYOE 40

DB 601 LNLNNTYYALKRGASTRSELPKNGSENENMLYMGKTS 640

RESULT 7

A41859

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influe

C:Species: Haemophilus influenzae

A:Variety: strain HK715

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: A41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influe

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: A41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1702 <POU>

A:Cross-references: GB:M87489; NID:g148906; PIDN:AAA24966.1; PID:g148907

A:Experimental source: strain HK715

A:Note: sequence extracted from NCBI backbone (NCBIP:97282)

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 41.1%; Score 92; DB 2; Length 1702;

Best Local Similarity 37.5%; Pred. No. 0.0076;

Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYKRYRYALKSGRLNMPENGVAENNDVWFGTYOE 40

DB 601 LNLNNTYYALKRGASTRSELPKNGSENENMLYMGKTS 640

RESULT 8

C41859

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK613

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: C41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influe

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: C41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613

A:Note: sequence extracted from NCBI backbone (NCBIP:97285)

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 37.5%; Score 84; DB 2; Length 1849;

Best Local Similarity 35.7%; Pred. No. 0.01;

Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 LYK--NYRYALKSGRLNMPENGVAENNDVWFGTYOE 40

DB 604 LYFNODRSTYTLKGASTRSELPONSGSENENMLYMGRTSD 645

RESULT 9

B41859

IgA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK393

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: B41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influe

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: B41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1545 <POU>

A:Cross-references: GB:M87490; NID:g148908; PIDN:AAA24967.1; PID:g148909

A:Experimental source: strain HK393

A:Note: sequence extracted from NCBI backbone (NCBIP:97283)

C:Superfamily: IgA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloproteinase

Query Match 36.2%; Score 81; DB 2; Length 1545;

Best Local Similarity 39.5%; Pred. No. 0.021;

Matches 15; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 1 LY--KNRYRYALKSGRLNMPENGVAENNDVWFG 36

DB 597 LYFNEENRTYALKKDAISRSEFPONRGSNNMSWLYMG 634

RESULT 10

CBRY

carboxypeptidase C (EC 3.4.16.5) precursor [validated] - yeast (Saccharomyces cerevis

M:Alternate names: carboxypeptidase Y; protein YKR297w

C:Species: Saccharomyces cerevisiae

C:Date: 15-Oct-1982 #sequence_revision 31-Mar-1993 #text_change 15-Sep-2000

C:Accession: A26597; S47458; A90763; A94609; A00909

R:Valls, L.A.; Hunter, C.P.; Rothman, J.H.; Stevens, T.H.

Cell 48, 887-897, 1987

A:Title: Protein sorting in yeast: the localization determinant of yeast vacuolar car

A:Reference number: A26597; MUID:87131100; PMID:3028649

A:Accession: A26597

RESULT 14

AD517
UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (EC C:Species: *Salmonella enterica* subsp. *enterica* serovar Typh
A:Note: this species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 03-Jun-2002
C:Accession: A10517
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, th, T.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serov A:Reference number: AB0502; PMID:11677608
A:Accession: A10517
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD01281.1; PID:g16501409; GSPDB:GN00176
C:Genetics:
A:Gene: SPY0144
C:Keywords: ligase

Query Match 27.2%; Score 61; DB 2; Length 452;
Best Local Similarity 42.3%; Pred. No. 2.6;
Matches 11; Conservative 6; Mismatches 5; Indels 4; Gaps 1;

OY 12 KSGRLNAPMPENG---AENNDWV 33

Db 200 KAKGEITGTGPIENGAIAMNADNDWL 225

RESULT 15

F64730

UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diamino-pimelate-D-alanyl-D-alanine ligase (EC N:Alternate names: UDP-N-acetylmuramoyl-pentapeptide synthetase
C:Species: *Escherichia coli*
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 03-Jun-2002
C:Accession: F64730; S04846; S40596
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; CC A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of *Escherichia coli* K-12.
A:Reference number: A64720; M0ID:97426617; PMID:9278503
A:Accession: F64730
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <BLAT>
A:Cross-references: GB:AE000118; GB:U00096; M0D:g1786262; PIDN:AACT3197.1; PID:g1786274; A:Experimental source: strain K-12, substrain MG1655
R:Parquet, C.; Plouret, B.; Mengin-Lecreulx, D.; van Heijenoort, J.
Nucleic Acids Res. 17, 5379, 1989
A:Title: Nucleotide sequence of the murF gene encoding the UDP-murNAc-pentapeptide synth A:Reference number: S04846; M0ID:89345095; PMID:2668880
A:Accession: S04846
A:Molecule type: DNA
A:Residues: 1-60, 'A', 62-177, 'R', 179-452 <PAR>
A:Cross-references: EMBL:X15432; M0D:g42047; PIDN:CAA33473.1; PID:g42048
R:Experimental source: strain K-12
R:Yura, T.; Mori, H.; Nagai, H.; Nagata, T.; Ishihama, A.; Fujita, N.; Isono, K.; Mizobu submitted to the EMBL Data Library, December 1992
A:Description: Systematic sequencing of the *Escherichia coli* genome: analysis of the 0-2 A:Reference number: S40531
A:Accession: S40596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-60, 'A', 62-177, 'R', 179-452 <YUR>
A:Cross-references: EMBL:D10483; M0D:g216434; PIDN:BA01351.1; PID:d1001823; PID:g216500 C:Genetics:
A:Gene: murF
A:Map position: 2 min
A:Function:
A:Pathway: peptidoglycan biosynthesis

C:Keywords: ATP; cell division; cell wall; ligase; monomer; nucleotide binding; P-loop
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-452/Product: UDP-N-acetylmuramoylalanyl-D-glutamyl-2,6-diaminopimelate-D-alanyl- F:107-112/Region: nucleotide-binding motif A (P-loop)

Query Match 27.2%; Score 61; DB 2; Length 452;
Best Local Similarity 42.3%; Pred. No. 2.6;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;

OY 12 KSGRLNAPMPENG---AENNDWV 33

Db 200 KAKGEITGTGPIENGAIAMNADNDWL 225

Search completed: December 10, 2002, 10:57:41
Job time : 13 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:17:38 ; Search time 5.6 Seconds

(without alignments)
296.259 Million cell updates/sec

Title: US-09-142-970-4_COPY_1_40
Perfect score: 224
Sequence: 1 LYKKNRYALKSGRLNAPMPENGVAENNDVWVGWYQTE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	223	99.6	1532	1 IGA_NEIGO	P09790 neisseria g
2	92	41.1	1541	1 IGA1_HAEIN	P42782 haemophilus
3	92	41.1	1694	1 IGA0_HAEIN	P44969 haemophilus
4	92	41.1	1702	1 IGA2_HAEIN	P45384 haemophilus
5	84	37.5	1849	1 IGA4_HAEIN	P45385 haemophilus
6	81	36.2	1545	1 IGA3_HAEIN	P45386 haemophilus
7	62.5	27.9	532	1 CBPY_YEAST	P00729 saccharomyc
8	62	27.7	297	1 YMY9_YEAST	Q03161 saccharomyc
9	61	27.2	452	1 MURF_ECOLI	P19137 mus musculu
10	58	25.9	3084	1 LMA1_MOUSE	P19137 mus musculu
11	56	25.0	3075	1 TSH_DROME	P22265 drosophila
12	55.5	24.8	993	1 RNT2_YEAST	P02467 streptococ
13	54.5	23.9	658	1 RNT2_YEAST	P32843 saccharomyc
14	53.5	23.7	1076	1 CARB_ANCFU	O28994 archaeoglob
15	53	23.4	671	1 ALYS_ENTFA	P37710 enterococcu
16	52.5	23.2	443	1 YME3_YEAST	P39959 saccharomyc
17	52	23.0	507	1 YME3_YEAST	O04712 saccharomyc
18	51.5	23.0	4967	1 RYR2_HUMAN	Q92736 homo sapien
19	51.5	23.0	4969	1 RYR2_HUMAN	Q92736 homo sapien
20	51.5	23.0	4969	1 RYR2_HUMAN	Q92736 homo sapien
21	51	22.8	1193	1 DP3A_XYLF	O99047 xylella fas
22	50.5	22.5	378	1 RFBQ_SALBO	P52442 salmonella
23	50.5	22.5	1361	1 RPOD_SPIOLO	P11704 spinacia ol
24	50	22.3	359	1 RYH7_HUMAN	P21854 homo sapien
25	50	22.3	641	1 RYH7_HUMAN	P21854 homo sapien
26	50	22.3	641	1 RYH7_HUMAN	P21854 homo sapien
27	50	22.3	801	1 RYH7_HUMAN	P21854 homo sapien
28	49.5	22.1	114	1 VORA_PYRHO	O58413 pyrococcus
29	49	21.9	588	1 BIN1_MOUSE	O08839 mus musculu
30	49	21.9	588	1 BIN1_MOUSE	O08839 mus musculu
31	49	21.9	588	1 BIN1_MOUSE	O08839 mus musculu
32	49	21.9	588	1 BIN1_MOUSE	O08839 mus musculu
33	48.5	21.7	361	1 SEP2_HUMAN	O15019 homo sapien

34	48.5	21.7	361	1 SEP2_MOUSE	P42208 mus musculu
35	48.5	21.7	520	1 Y120_MYCGE	P47366 mycoplasma
36	48.5	21.7	533	1 NIPD_CLOPA	P00467 clostridium
37	48.5	21.7	704	1 GYS2_YEAST	P27472 saccharomyc
38	48.5	21.7	711	1 J1PI_HUMAN	O94612 homo sapien
39	48.5	21.7	1225	1 YF78_MYCGE	P47551 mycoplasma
40	48	21.4	100	1 YF78_MYCGE	P75202 mycoplasma
41	48	21.4	106	1 Y086_CAEEL	O09238 caenorhabdi
42	48	21.4	291	1 Y32K_SSV1	P20197 sulfolobus
43	48	21.4	296	1 ALYS_BDPPI	O03979 bacterioph
44	48	21.4	366	1 YF59_STAMM	O09177 staphylococ
45	48	21.4	1176	1 CIAA_BACTK	P02965 bacillus th

ALIGNMENTS

RESULT 1
IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RC MEDLINE=87115823; PubMed=3027577;
RA Fohlner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT Iga protease.";
RL Nature 325:458-462(1987).
RP
CC ACTIVE SITE.
RA MEDLINE=90154052; PubMed=2105953;
RA Bachovchin W.W., Plant A.G., Flentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of Iga1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide boronic acids.";
RL J. Biol. Chem. 265:3738-3743(1990).
CC FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X04835; CAA28538.1; -
CC PIR: A26039; A26039.
CC MEROPS: S06.001; -
CC InterPro: IPR000710; IGA-S6.
CC InterPro: IPR004899; Peptidase-sup.
CC Pfam: PF02395; IGA1; 1.
CC Pfam: PF03212; Peptidase; 1.
CC PRINTS: PRO0921; IGASERPTASE.
CC Hydrolyase; Serine protease; zymogen; Autocatalytic cleavage;

```

KW Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 28 96 IGA-SPECIFIC SPRINE ENDOPEPTIDASE.
FT PROPEP 987 1532 HELPER PEPTIDE.
FT ACT_SITE 278 278 POTENTIAL.
FT SITE 986 987 CLEAVAGE (AUTO-).
FT SITE 1018 1019 CLEAVAGE (AUTO-).
FT SITE 1121 1122 CLEAVAGE (AUTO-).
SQ SEQUENCE 1532 AA; 168976 MW; 68F4112BD22F40D CRC64;

Query Match 99.6%; Score 223; DB 1; Length 1532;
Best Local Similarity 97.5%; Pred. No. 1.6e-22;
Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGRLNAPMPENGVAENNDWFMGYTQE 40
DB 584 LYKKNRYVALKSGRLNAPMPENGVAENNDWFMGYTQE 623

RESULT 2
ID IGA1_HAEIN STANDARD; PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OC NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK368 / Serotype B;
RA MEDLINE=69379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Kilian M.;
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT of Haemophilus influenzae/serotype b.";
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SER-288.
RC STRAIN-HK368 / Serotype B;
RA MEDLINE=92324949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Kilian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RL Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X64357; CAA45708.1;
CC DR EMBL: M87482; AAA24969.1;
CC DR MEROPS: S06.001;
CC DR InterPro: IPR000710; IGA_S6.

```

```

DR InterPro: IPR004899; Pertact-sup.
DR Pfam: PF02395; IGA1_1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KW Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT MUTAGEN 288 288 S->T: LOSS OF ACTIVITY.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

Query Match 41.1%; Score 92; DB 1; Length 1541;
Best Local Similarity 37.5%; Pred. No. 0.00018;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGRLNAPMPENGVAENNDWFMGYTQE 40
DB 595 LNKKNRYVALKSGRLNAPMPENGVAENNDWFMGYTQE 634

RESULT 3
ID IGA0_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX Haemophilus.
OC NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Serotype D;
RA Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RA MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA McKelvey K., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddiom E., Cotton M.D.,
RA Uiterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fleischmann J., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.E., McDonald L.A., Shail K.V., Fraser C.M., Smith H.O.,
RA Verter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: VIRULENCE FACTOR. CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X59800; NOT ANNOTATED_CDS.
CC EMBL: U32779; AAC22651.1; .
CC MEROPS: S06.001; .
CC TIGR: H10990; .
CC InterPro: IPR000710; IGA-S6.
CC InterPro: IPR004899; Pertactin_sup.
CC Pfam: PF02395; IGA1; 1.
CC Pfam: PF03212; Pertactin; 2.
CC PRINTS: PR00921; IGASERPRASE.
CC HydroLase; Serine protease; Transmembrane; zymogen; Signal;
CC Complete proteome.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
CC PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
CC ACT_SITE 288 288 PROBABLE.
CC EN -> GV (IN REF. 1).
CC FT CONFLICT 272 272 G -> A (IN REF. 1).
CC FT CONFLICT 464 464 G -> E (IN REF. 1).
CC FT CONFLICT 866 866 S -> T (IN REF. 1).
CC FT CONFLICT 1036 1036 A -> D (IN REF. 1).
CC FT CONFLICT 1074 1074 A -> G (IN REF. 1).
CC FT CONFLICT 1421 1421 A -> G (IN REF. 1).
CC FT CONFLICT 1545 1545 H -> T (IN REF. 1).
CC SEQUENCE 1694 AA; 185539 MW; C52427013F93176C CRC64;
SQ

Query Match 41.1%; Score 92; DB 1; Length 1694;
Best Local Similarity 37.5%; Pred. No. 0.0002;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGGRINAPMPENGAENNDAWVFGYTOE 40
DB 601 LNEENTYVALRKGASTRSELPKNSGSENNWLYMGKTS 640

RESULT 4
IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RA "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RT Influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC substrates are known.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M87489; AAA24966.1; .
CC MEROPS: S06.001; .
CC InterPro: IPR000710; IGA-S6.
CC InterPro: IPR004899; Pertactin_sup.
CC Pfam: PF02395; IGA1; 1.
CC Pfam: PF03212; Pertactin; 2.
CC PRINTS: PR00921; IGASERPRASE.
CC HydroLase; Serine protease; Transmembrane; zymogen; Repeat; Signal;
CC Complete proteome.
CC SIGNAL 1 25 POTENTIAL.
CC CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
CC PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
CC ACT_SITE 288 288 PROBABLE.
CC DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
CC FT REPEAT 1109 1116 K.
CC FT REPEAT 1117 1124 2.
CC SEQUENCE 1702 AA; 186539 MW; 860F70D2667807A6 CRC64;
SQ

Query Match 41.1%; Score 92; DB 1; Length 1702;
Best Local Similarity 37.5%; Pred. No. 0.0002;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGGRINAPMPENGAENNDAWVFGYTOE 40
DB 601 LNEENTYVALRKGASTRSELPKNSGSENNWLYMGKTS 640

RESULT 5
IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NHTI HK61;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RA "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@slb.ch.
CC -----
DR EMBL: M87491; AAA24968.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
KM Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1021 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1022 1849 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 299 299 PROBABLE.
SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 37.5%; Score 84; DB 1; Length 1849;
Best Local Similarity 35.7%; Pred. No. 0.0028;
Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 LYYK--NRYRYALKSGRLNAPMPENGVAENNDVFMGYTQE 40
Db 604 LYFNQDNRSYTYLKKGASTSELPQNGESNENMLYMGRTSD 645
||: ||||| :||| :||:| |:

RESULT 6
IGA3_HAEIN STANDARD; PRT; 1545 AA.
ID IGA3_HAEIN
AC P45385;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_TaxId=727;
ON [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HK393 / NCTC 8467 / Serotype B;
RX MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
influenzae type 1 immunoglobulin A1 proteases.";
RL J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A
PRODUCING INTACT FC AND FAB FRAGMENTS.
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
certain Pro-I-Xaa bonds in the hinge region. No small molecule
substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
SPACE, THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@slb.ch).
CC -----
DR EMBL: M87490; AAA24967.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.

DR PRINTS: PR00921; IGASERPTASE.
KM Hydrolyase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 292 292 PROBABLE.
SQ SEQUENCE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;

Query Match 36.2%; Score 81; DB 1; Length 1545;
Best Local Similarity 39.5%; Pred. No. 0.0058;
Matches 15; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 1 LYY--KNRYRYALKSGRLNAPMPENGVAENNDVFMG 36
Db 597 LYFNEENRITYYALKKASIRSEPPQNGESNENMLYMG 634
||: : ||||| :||| :||:| |:

RESULT 7
CBPY_YEAST STANDARD; PRT; 532 AA.
ID CBPY_YEAST
AC P00729;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).
GN PC1 OR YMR297W.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
NCBI_TaxId=4932;
ON [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87131100; PubMed=3028649;
RA Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;
RT "Protein sorting in yeast: the localization determinant of yeast
vacuolar carboxypeptidase Y resides in the propeptide.";
RL Cell 48:887-897(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Badcock K., Churcher C., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE OF 112-532.
RA Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;
RT "Amino acid sequence of carboxypeptidase Y. II. Peptides from
enzymatic cleavages.";
RL Carlsberg Res. Commun. 47:15-27(1982).
RN [4]
RP REVISIONS, AND ACTIVE SITE SER-257.
RA Bredam K., Svendsen I.;
RT "Identification of methionyl and cysteinyl residues in the substrate
binding site of carboxypeptidase Y.";
RL Carlsberg Res. Commun. 49:639-645(1984).
RN [5]
RP ACTIVE SITE HIS-508.
RX MEDLINE=90315013; PubMed=2639680;
RA Bech L.M., Bredam K.;
RT "Inactivation of carboxypeptidase Y by mutational removal of the
putative essential histidyl residue.";
RL Carlsberg Res. Commun. 54:165-171(1989).
RN [6]
RP MUTAGENESIS.
RX MEDLINE=94114535; PubMed=7904479;
RA Mortensen U.H., Remington S.J., Bredam K.;
RT "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen
bond network stabilizes the transition state by interaction with the
C-terminal carboxylate group of the substrate.";
RL Biochemistry 33:508-517(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
RX MEDLINE=95244421; PubMed=7727362;
RA Endritzl J.A., Bredam K., Remington S.J.;

RA Parquet C., Flouret B., Mengin-Lecreux D., van Heijenoort J.;
 RT "Nucleotide sequence of the murf gene encoding the UDP-MurNAc-
 RT pentapeptide synthetase of *Escherichia coli*.";
 RL Nucleic Acids Res. 17:5379-5379(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=92334977; PubMed=1630901;
 RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
 RA Isono K., Mizobuchi K., Nakata A.;
 RT "Systematic sequencing of the *Escherichia coli* genome: analysis of
 RT the 0-2.4 min region.";
 RL Nucleic Acids Res. 20:3305-3308(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MB2884;
 RX MEDLINE=97128642; PubMed=8973200;
 RA Anderson M.S., Eveland S.S., Onishi H.R., Pompliano D.L.;
 RT "Kinetic mechanism of the *Escherichia coli* UDPmurNac-tripeptide
 RT gamma-D-alanine-adding enzyme: use of a glutathione S-transferase
 RT fusion.";
 RL Biochemistry 35:16264-16269(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:12453-12474(1997).
 RN [5]
 RP SEQUENCE FROM N.A., AND MUTANT MURF2.
 RC STRAIN-CGSC 5990;
 RX MEDLINE=97309380; PubMed=9166795;
 RA Eveland S.S., Pompliano D.L., Anderson M.S.;
 RT "Conditionally lethal *Escherichia coli* murE mutants contain point
 RT defects that map to regions conserved among murein and folyl poly-
 RT gamma-glutamate ligases: identification of a ligase superfamily.";
 RL Biochemistry 36:6223-6229(1997).
 RN [6]
 RP CHARACTERIZATION, AND SEQUENCE OF 1-15.
 RX MEDLINE=90248455; PubMed=2168811;
 RA Duncan K., van Heijenoort J., Walsh C.T.;
 RT "Purification and characterization of the D-alanyl-D-alanine-adding
 RT enzyme from *Escherichia coli*.";
 RL Biochemistry 29:2379-2386(1990).
 RN [7]
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
 RX MEDLINE=20545602; PubMed=11090285;
 RA Yan Y., Munshi S., Leiting B., Anderson M.S., Chiras J., Chen Z.;
 RT "Crystal structure of *Escherichia coli* UDPmurNac-tripeptide
 RT D-alanyl-D-alanine-adding enzyme (MurF) at 2.3-A resolution.";
 RL J. Mol. Biol. 304:435-445(2000).
 CC - FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
 CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLURAMOYL-PENTAPEPTIDE, THE
 CC PRECURSOR OF MUREIN.
 CC - CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
 CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
 CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
 CC carboxy-L-lysyl-D-alanyl-D-alanine.
 CC - PATHWAY: Peptidoglycan biosynthesis.
 CC - SUBUNIT: MONOMER.
 CC - SUBCELLULAR LOCATION: Cytoplasmic.
 CC - SIMILARITY: BELONGS TO THE MURCER FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>).

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL: X15432; CA33473.1; -
 DR EMBL: X55034; CA38863.1; -
 DR EMBL: D10483; BA01351.1; -
 DR EMBL: U67891; AAC44657.1; -
 DR EMBL: AE000118; AAC73197.1; -
 DR EMBL: U67893; AAB60788.1; -
 DR PIR: S04846; S04846.
 DR PIR: S40596; S40596.
 DR PDB: 1GG4; 20-DEC-00.
 DR Ecogene: EG10622; murF.
 DR InterPro: IPR000713; Mur_ligase.
 DR InterPro: IPR004101; Mur_ligase_C.
 DR Pfam: PF01225; Mur_ligase_1.
 DR Pfam: PF02875; Mur_ligase_C_1.
 DR TIGRfams: TIGR01143; murF_1.
 KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
 KW ATP-binding; 3D-structure; Complete proteome.
 FT NP_BIND 107 113 ATP (POTENTIAL).
 FT VARIANT 288 288 A -> T (IN MURF2; TS MUTANT WITH LOW
 FT ACTIVITY).
 FT CONFLICT 61 61 G -> A (IN REF. 1 AND 2).
 FT CONFLICT 178 178 A -> R (IN REF. 1 AND 2).
 SQ SEQUENCE 452 AA; 47447 MW; B46E2E57BDB8C572 CRC64;
 Query Match 27.2%; Score 61; DB 1; Length 452;
 Best Local Similarity 42.3%; Pred. No. 0.81;
 Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;
 Oy 12 KSGRLNAMPENGV---AEENDMV 33
 Db 200 KAKGEIFSGLPENGIAIMNADNDWL 225
 ID LMA1_MOUSE STANDARD; PRT; 3084 AA.
 AC P19137;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lamnin alpha-1 chain precursor (Lamnin A chain).
 GN LAMA1 OR LAMA-1 OR LAMA.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89034134; PubMed=3182802;
 RA Sasaki M., Kleinman H.K., Huber H., Deutzmann R., Yamada Y.;
 RT "Lamnin, a multidomain protein. The A chain has a unique globular
 RT domain and homology with the basement membrane proteoglycan and the
 RT lamnin B chains.";
 RL J. Biol. Chem. 263:16536-16544(1988).
 RN [2]
 RP SEQUENCE OF 1-339 FROM N.A.
 RX MEDLINE=88225080; PubMed=3267223;
 RA Hartl L., Oberbauer I., Deutzmann R.;
 RT "The N terminus of lamnin A chain is homologous to the B chains.";
 RL Eur. J. Biochem. 173:629-635(1988).
 RN [3]
 RP SEQUENCE OF 2538-3084 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=89030693; PubMed=3181157;
 RA Deutzmann R., Huber J., Schmetz K.A., Oberbauer I., Hartl L.;
 RT "Structural study of long arm fragments of lamnin. Evidence for
 RT repetitive C-terminal sequences in the A-chain, not present in the B-
 RT chains.";
 RL Eur. J. Biochem. 177:35-45(1988).
 CC - FUNCTION: Binding to cells via a high affinity receptor, lamnin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting

CC with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
CC LAMININ-3 (S-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI, IV AND G ARE GLOBULAR.
CC -1- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -1- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: J04064; AAA39410.1; -
CC EMBL: X07737; CA30561.1; -
CC EMBL: X13459; CA31807.1; -
CC EMBL: M36775; AAA39406.1; -
CC PIR: A31771; MMSA.
CC HSSP: O60675; IOUO.
CC MGD: MG:99892; Lam1.
CC InterPro: IPR000561; EGF-like.
CC InterPro: IPR001886; LamNT.
CC InterPro: IPR000034; Laminin_B.
CC InterPro: IPR002049; Laminin_EGF.
CC InterPro: IPR001791; Laminin_G.
CC Pfam: PF00052; Laminin_B_2.
CC Pfam: PF00053; Laminin_EGF_15.
CC Pfam: PF00054; Laminin_G_5.
CC Pfam: PF00055; Laminin_Nterm_1.
CC PRINTS: PR00011; EGF/LAMININ.
CC ProDom: PD002082; LamNT_1.
CC ProDom: PD003031; Laminin_B_2.
CC SMART: SM00180; EGF_Lam; 14.
CC SMART: SM00281; LamB; 2.
CC SMART: SM00282; LamG; 5.
CC SMART: SM00136; LamNT; 1.
CC PROSITE: PS00022; EGF_1; 11.
CC PROSITE: PS01186; EGF_2; 3.
CC PROSITE: PS01248; LAMININ_TYPE_EGF; 15.
CC PROSITE: PS50025; LAM_G_DOMAIN; 5.
CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
CC FT SIGNAL 1 24 POTENTIAL.
CC FT CHAIN 25 3084 LAMININ ALPHA-1 CHAIN.
CC FT MOD_RES 25 25 BLOCKED.
FT DOMAIN 25 276 LAMININ N-TERMINAL (DOMAIN VI).
FT DOMAIN 277 333 LAMININ EGF-LIKE 1.
FT DOMAIN 334 403 LAMININ EGF-LIKE 2.
FT DOMAIN 404 460 LAMININ EGF-LIKE 3.
FT DOMAIN 461 509 LAMININ EGF-LIKE 4.
FT DOMAIN 510 519 LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN 520 715 LAMININ DOMAIN IV 1 (DOMAIN IV B).
FT DOMAIN 716 748 LAMININ EGF-LIKE 6.
FT DOMAIN 749 797 LAMININ EGF-LIKE 5 (C-TERMINAL).
FT DOMAIN 798 855 LAMININ EGF-LIKE 7.
FT DOMAIN 856 908 LAMININ EGF-LIKE 8.
FT DOMAIN 909 957 LAMININ EGF-LIKE 9.
FT DOMAIN 958 1004 LAMININ EGF-LIKE 10.
FT DOMAIN 1005 1050 LAMININ EGF-LIKE 11.

FT DOMAIN 1051 1096 LAMININ EGF-LIKE 12.
FT DOMAIN 1097 1156 LAMININ EGF-LIKE 13.
FT DOMAIN 1157 1166 LAMININ EGF-LIKE 14 (N-TERMINAL).
FT DOMAIN 1167 1368 LAMININ DOMAIN IV 2 (DOMAIN IV A).
FT DOMAIN 1369 1409 LAMININ EGF-LIKE 14 (C-TERMINAL).
FT DOMAIN 1410 1458 LAMININ EGF-LIKE 15.
FT DOMAIN 1459 1515 LAMININ EGF-LIKE 16.
FT DOMAIN 1516 1562 LAMININ EGF-LIKE 17.
FT DOMAIN 1564 2124 LAMININ IT AND 1.
FT DOMAIN 2125 2305 LAMININ G-LIKE 1.
FT DOMAIN 2313 2489 LAMININ G-LIKE 2.
FT DOMAIN 2494 2680 LAMININ G-LIKE 3.
FT DOMAIN 2722 2894 LAMININ G-LIKE 4.
FT DOMAIN 2899 3079 LAMININ G-LIKE 5.
FT DOMAIN 1612 1820 COILED COIL (POTENTIAL).
FT DOMAIN 1869 1903 COILED COIL (POTENTIAL).
FT DOMAIN 2096 2128 COILED COIL (POTENTIAL).
FT DOMAIN 1147 1149 CELL ATTACHMENT SITE.
FT SITE 1147 1149
FT DISULFID 277 286 BY SIMILARITY.
FT DISULFID 279 297 BY SIMILARITY.
FT DISULFID 299 308 BY SIMILARITY.
FT DISULFID 311 331 BY SIMILARITY.
FT DISULFID 334 343 BY SIMILARITY.
FT DISULFID 336 368 BY SIMILARITY.
FT DISULFID 371 380 BY SIMILARITY.
FT DISULFID 383 401 BY SIMILARITY.
FT DISULFID 404 416 BY SIMILARITY.
FT DISULFID 406 434 BY SIMILARITY.
FT DISULFID 436 445 BY SIMILARITY.
FT DISULFID 448 458 BY SIMILARITY.
FT DISULFID 461 474 BY SIMILARITY.
FT DISULFID 463 478 BY SIMILARITY.
FT DISULFID 480 489 BY SIMILARITY.
FT DISULFID 492 507 BY SIMILARITY.
FT DISULFID 507 558 BY SIMILARITY.
FT DISULFID 749 758 BY SIMILARITY.
FT DISULFID 751 764 BY SIMILARITY.
FT DISULFID 767 776 BY SIMILARITY.
FT DISULFID 779 795 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 800 823 BY SIMILARITY.
FT DISULFID 826 835 BY SIMILARITY.
FT DISULFID 838 853 BY SIMILARITY.
FT DISULFID 856 870 BY SIMILARITY.
FT DISULFID 858 877 BY SIMILARITY.
FT DISULFID 880 889 BY SIMILARITY.
FT DISULFID 892 906 BY SIMILARITY.
FT DISULFID 909 921 BY SIMILARITY.
FT DISULFID 911 928 BY SIMILARITY.
FT DISULFID 930 939 BY SIMILARITY.
FT DISULFID 942 955 BY SIMILARITY.
FT DISULFID 958 970 BY SIMILARITY.
FT DISULFID 960 976 BY SIMILARITY.
FT DISULFID 978 987 BY SIMILARITY.
FT DISULFID 990 1002 BY SIMILARITY.
FT DISULFID 1005 1021 BY SIMILARITY.
FT DISULFID 1007 1014 BY SIMILARITY.
FT DISULFID 1023 1032 BY SIMILARITY.
FT DISULFID 1035 1048 BY SIMILARITY.
FT DISULFID 1051 1063 BY SIMILARITY.
FT DISULFID 1053 1070 BY SIMILARITY.
FT DISULFID 1072 1081 BY SIMILARITY.
FT DISULFID 1084 1094 BY SIMILARITY.
FT DISULFID 1410 1419 BY SIMILARITY.
FT DISULFID 1412 1426 BY SIMILARITY.
FT DISULFID 1429 1438 BY SIMILARITY.
FT DISULFID 1441 1456 BY SIMILARITY.
FT DISULFID 1459 1473 BY SIMILARITY.
FT DISULFID 1461 1483 BY SIMILARITY.
FT DISULFID 1486 1495 BY SIMILARITY.
FT DISULFID 1498 1513 BY SIMILARITY.
FT DISULFID 1516 1528 BY SIMILARITY.
FT DISULFID 1518 1535 BY SIMILARITY.
FT DISULFID 1537 1546 BY SIMILARITY.

FT DISULFID 1549 1560 BY SIMILARITY.
 FT DISULFID 1563 1563 INTERCHAIN (PROBABLE).
 FT DISULFID 1567 1567 INTERCHAIN (PROBABLE).
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 959 959 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1344 1344 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1414 1414 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 25.98; Score 58; DB 1; Length 3084;
 Best Local Similarity 30.0%; Pred. No. 19;
 Matches 15; Conservative 6; Mismatches 15; Indels 14; Gaps 2;

OY 5 NYRYVALKSGCR-----LNAPMPENGVAEEND-----WYMKGYTQE 40
 DB 1260 NYEPQYLKGGARKHYIYMDAPAPENGVRQDIYOMKEEFKXFNVSSE 1309

RESULT 11
 LMA1_HUMAN STANDARD; PRT; 3075 AA.
 AC P23391;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Laminin alpha-1 chain precursor (Laminin A chain).
 GN LAMA1 OR LAMA.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91333420; PubMed-1714537;
 RA Haaparanta T., Uitto J., Ruoslahti E., Engvall E.;
 RT "Molecular cloning of the cDNA encoding human laminin A chain.";
 RL Matrix 11:151-160(1991).
 RN [12]
 RP SEQUENCE OF 1-2628 FROM N.A.
 RX MEDLINE-91264789; PubMed-2049067;
 RA Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,
 RA Tryggvason K.;
 RT "Primary structure of the human laminin A chain. Limited expression
 in human tissues.";
 RL Biochem. J. 276:369-379(1991).
 RN [3]
 RP SEQUENCE OF 2397-3072 FROM N.A.
 RX MEDLINE-89280632; PubMed-2733383;
 RA Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
 RA Sanborn D., Sasaki T., Kuivaniemi H., Chu M.L., Dettmann R.,
 RA Timp R., Uitto J.;
 RT "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1
 and B2 chains, and expression of the corresponding genes in human
 skin and cultured cells.";
 RL Lab. Invest. 60:772-782(1989).
 CC -!- FUNCTION: Binding to cells via a high affinity receptor, laminin
 is thought to mediate the attachment, migration, and organization
 of cells into tissues during embryonic development by interacting
 with other extracellular matrix components.
 CC -!- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 different polypeptide chains (alpha, beta, gamma), which are bound
 to each other by disulfide bonds into a cross-shaped molecule
 comprising one long and three short arms with globules at each

CC end.
 CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
 CC LAMININ-3 (S-LAMININ).
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -!- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -!- DOMAIN: DOMAINS VI, IV AND G ARE GLOBALAR.
 CC -!- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
 CC -!- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
 CC -!- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X58531; CAA1418.1; -
 CC PIR: S14458; S14458.
 CC HSSP: Q0675; 1000.
 CC Genew: HGNC:6481; LAMA1.
 CC MIM: 150320; -
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR001886; LamNT.
 CC InterPro: IPR000034; Laminin_B.
 CC InterPro: IPR002049; Laminin_EGF.
 CC InterPro: IPR001791; Laminin_G.
 CC Pfam: PF00052; Laminin_B; 2.
 CC Pfam: PF00053; Laminin_EGF; 15.
 CC Pfam: PF00054; Laminin_G; 5.
 CC Pfam: PF00055; Laminin_NTerm; 1.
 CC PRINTS: PR00011; EGF/LAMININ.
 CC ProDom: PD002082; LamNT; 1.
 CC ProDom: PD003031; Laminin_B; 2.
 CC SMART: SM00180; EGF_Lam; 14.
 CC SMART: SM00001; EGF_Like; 1.
 CC SMART: SM00281; Lamb; 2.
 CC SMART: SM00282; Lamb; 5.
 CC SMART: SM00136; LamNT; 1.
 CC PROSITE: PS00186; EGF_1; 11.
 CC PROSITE: PS00182; EGF_2; 2.
 CC PROSITE: PS01248; LAMININ_TYPE_EGF; 15.
 CC PROSITE: PS50025; LAM_G_DOMAIN; 5.
 CC Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
 CC Laminin EGF-like domain; Cell adhesion; Repeat; Signal.
 CC SIGNAL 1 17
 CC FT CHAIN 18 3075 LAMININ ALPHA-1 CHAIN.
 CC FT CHAIN 18 269 LAMININ N-TERMINAL (DOMAIN VI).
 CC FT DOMAIN 270 326 LAMININ EGF-LIKE 1.
 CC FT DOMAIN 327 396 LAMININ EGF-LIKE 2.
 CC FT DOMAIN 397 453 LAMININ EGF-LIKE 3.
 CC FT DOMAIN 454 502 LAMININ EGF-LIKE 4.
 CC FT DOMAIN 503 512 LAMININ EGF-LIKE 5 (N-TERMINAL).
 CC FT DOMAIN 517 708 LAMININ EGF-LIKE 14 (DOMAIN IV B).
 CC FT DOMAIN 709 741 LAMININ EGF-LIKE 5 (C-TERMINAL).
 CC FT DOMAIN 742 790 LAMININ EGF-LIKE 6.
 CC FT DOMAIN 791 848 LAMININ EGF-LIKE 7.
 CC FT DOMAIN 849 901 LAMININ EGF-LIKE 8.
 CC FT DOMAIN 902 950 LAMININ EGF-LIKE 9.
 CC FT DOMAIN 951 997 LAMININ EGF-LIKE 10.
 CC FT DOMAIN 998 1043 LAMININ EGF-LIKE 11.
 CC FT DOMAIN 1044 1089 LAMININ EGF-LIKE 12.
 CC FT DOMAIN 1090 1149 LAMININ EGF-LIKE 13.
 CC FT DOMAIN 1150 1159 LAMININ EGF-LIKE 14 (N-TERMINAL).
 CC FT DOMAIN 1160 1361 LAMININ EGF-LIKE 2 (DOMAIN IV A).
 CC FT DOMAIN 1362 1402 LAMININ EGF-LIKE 14 (C-TERMINAL).
 CC FT DOMAIN 1403 1451 LAMININ EGF-LIKE 15.
 CC FT DOMAIN 1452 1508 LAMININ EGF-LIKE 16.

```

FT DOMAIN 1509 1555 LAMININ EGF-LIKE 17.
FT DOMAIN 1556 2116 DOMAIN II AND I.
FT DOMAIN 2117 2297 LAMININ G-LIKE 1.
FT DOMAIN 2305 2481 LAMININ G-LIKE 2.
FT DOMAIN 2486 2673 LAMININ G-LIKE 3.
FT DOMAIN 2713 2885 LAMININ G-LIKE 4.
FT DOMAIN 2890 3070 LAMININ G-LIKE 5.
FT DOMAIN 1706 1796 COILED COIL (POTENTIAL).
FT DOMAIN 1968 1989 COILED COIL (POTENTIAL).
FT DOMAIN 2088 2120 COILED COIL (POTENTIAL).
FT SITE 2534 2536 CELL ATTACHMENT SITE.
FT DISULFID 270 279 BY SIMILARITY.
FT DISULFID 272 290 BY SIMILARITY.
FT DISULFID 292 301 BY SIMILARITY.
FT DISULFID 297 305 POTENTIAL.
FT DISULFID 304 324 BY SIMILARITY.
FT DISULFID 327 336 BY SIMILARITY.
FT DISULFID 329 361 BY SIMILARITY.
FT DISULFID 364 373 BY SIMILARITY.
FT DISULFID 376 394 BY SIMILARITY.
FT DISULFID 397 409 BY SIMILARITY.
FT DISULFID 427 438 BY SIMILARITY.
FT DISULFID 429 438 BY SIMILARITY.
FT DISULFID 441 451 BY SIMILARITY.
FT DISULFID 454 467 BY SIMILARITY.
FT DISULFID 456 471 BY SIMILARITY.
FT DISULFID 473 482 BY SIMILARITY.
FT DISULFID 485 500 BY SIMILARITY.
FT DISULFID 742 751 BY SIMILARITY.
FT DISULFID 744 757 BY SIMILARITY.
FT DISULFID 760 769 BY SIMILARITY.
FT DISULFID 772 788 BY SIMILARITY.
FT DISULFID 791 806 BY SIMILARITY.
FT DISULFID 793 816 BY SIMILARITY.
FT DISULFID 819 828 BY SIMILARITY.
FT DISULFID 831 846 BY SIMILARITY.
FT DISULFID 849 863 BY SIMILARITY.
FT DISULFID 851 870 BY SIMILARITY.
FT DISULFID 873 882 BY SIMILARITY.
FT DISULFID 885 899 BY SIMILARITY.
FT DISULFID 902 914 BY SIMILARITY.
FT DISULFID 904 921 BY SIMILARITY.
FT DISULFID 923 932 BY SIMILARITY.
FT DISULFID 935 948 BY SIMILARITY.
FT DISULFID 951 963 BY SIMILARITY.
FT DISULFID 953 963 BY SIMILARITY.
FT DISULFID 971 980 BY SIMILARITY.
FT DISULFID 983 995 BY SIMILARITY.
FT DISULFID 998 1007 BY SIMILARITY.
FT DISULFID 1000 1014 BY SIMILARITY.
FT DISULFID 1016 1025 BY SIMILARITY.
FT DISULFID 1028 1041 BY SIMILARITY.
FT DISULFID 1044 1056 BY SIMILARITY.
FT DISULFID 1046 1063 BY SIMILARITY.
FT DISULFID 1065 1074 BY SIMILARITY.
FT DISULFID 1077 1087 BY SIMILARITY.
FT DISULFID 1403 1412 BY SIMILARITY.
FT DISULFID 1405 1419 BY SIMILARITY.
FT DISULFID 1422 1431 BY SIMILARITY.
FT DISULFID 1434 1449 BY SIMILARITY.
FT DISULFID 1452 1466 BY SIMILARITY.
FT DISULFID 1454 1476 BY SIMILARITY.
FT DISULFID 1479 1488 BY SIMILARITY.
FT DISULFID 1491 1506 BY SIMILARITY.
FT DISULFID 1509 1521 BY SIMILARITY.
FT DISULFID 1511 1528 BY SIMILARITY.
FT DISULFID 1530 1539 BY SIMILARITY.
FT DISULFID 1542 1553 BY SIMILARITY.
FT DISULFID 1556 1566 INTERCHAIN (PROBABLE).
FT DISULFID 1560 1560 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 555 555 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 763 763 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 926 926 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 952 952 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1407 1407 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1579 1579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1596 1596 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1678 1678 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1689 1689 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1698 1698 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1717 1717 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 25.0%; Score 56; DB 1; Length 3075;
Best Local Similarity 30.0%; Pred. No. 36;
Matches 15; Conservative 5; Mismatches 16; Indels 14; Gaps 2;

QY 5 NYRYALKSGR-----LNAPENGVAE-----NDWVFMGYTQE 40
Db 1253 NFEQVLIKGRIRKQYIMDAPAPENGVRQEQEVARENFMYKFSVSE 1302
: : ||| : || ||||| : :
TSH_DROME STANDARD; PRT; 993 AA.
ID TSH_DROME
AC P22265;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Teashirt protein.
TSH.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91098655; PubMed=1846092;
RA Fasano L., Roeder L., Core N., Alexandre E., Voia C., Jacq B.,
RA Kerridge S.;
RT "The gene teashirt is required for the development of Drosophila
RT embryonic trunk segments and encodes a protein with widely spaced
RT zinc finger motifs.";
RT Cell 64:63-79(1991).
RN [2]
RP POSSIBLE FUNCTION.
RX MEDLINE=93083418; PubMed=1360402;
RA Roeder L., Voia C., Kerridge S.;
RT "The role of the teashirt gene in trunk segmental identity in
RT Drosophila.";
RL Development 115:1017-1033(1992).
RN [3]
RP POSSIBLE FUNCTION.
RX MEDLINE=95009555; PubMed=7925029;
RA de Zulueta P., Alexandre E., Jacq B., Kerridge S.;
RT "Homeotic complex and teashirt genes co-operate to establish trunk
RT segmental identities in Drosophila.";
RL Development 120:2287-2296(1994).
CC -1- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF DROSOPHILA EMBRYONIC
CC TRUNK SEGMENTS. REPRESS THE EXPRESSION OF HEAD HOMOTIC GENES.
CC NECESSARY, IN COMBINATION WITH SCR, FOR THE FORMATION OF THE
CC PROTHORACIC SEGMENT. MAY BE INVOLVED IN CHROMATIN STRUCTURE FOR
CC MODULATION OF TRANSCRIPTION. NEGATIVELY CONTROL THE EXPRESSION OF
CC MOD AND POSITIVELY THAT OF DLT AND OF ITS OWN EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE TRUNK (FROM PS3
CC TO PS13).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL AND
CC ADULT DEVELOPMENT. NOT MATERNALLY EXPRESSED.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M57496; AAA28983.1; -.
DR PIR; A38437; A38437.
DR TRANSFAC; T00805; -.
DR FLYBase; FBgn0003866; tsn.
DR InterPro; IPR000823; Znf.C2H2.
DR Pfam; PF00096; zf-C2H2.3.
DR SMART; SM00355; ZNF_C2H2.3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS50157; ZINC_FINGER_C2H2_2; 3.
DR Developmental protein; Transcription regulation; Repressor; Activator;
KM Zinc-finger; Metal-binding; Repeat; DNA-binding; Nuclear protein.
FT DOMAIN 104 136 ALA-RICH
FT DOMAIN 175 183 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 354 357 ZINC FINGERS.
FT ZN_FING 354 378 C2H2-TYPE.
FT ZN_FING 466 490 C2H2-TYPE.
FT ZN_FING 533 557 C2H2-TYPE.
FT DOMAIN 104 107 POLY-ALA.
FT DOMAIN 115 122 POLY-ALA.
FT DOMAIN 175 180 POLY-GLU.
FT DOMAIN 401 407 POLY-PRO.
FT DOMAIN 830 834 POLY-ASN.
SQ SEQUENCE 993 AA; 106206 MW; 2DF9C6774F68B6D1 CRC64;

Query Match 24.8%; Score 55.5; DB 1; Length 993;
Best Local Similarity 38.7%; Pred. No. 11;
Matches 12; Conservative 5; Mismatches 5; Indels 9; Gaps 1;

OY 2 YKMYRYALKSGS-----RLNAPMPE 23
||:||||:|
DB 740 YQHYRTTSERSGSECAARPLDAPPE 770

RESULT 13
LYTB_STRPN STANDARD; PRT; 658 AA.
ID LYTB_STRPN
AC Q9Z4P;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.96)
DE (Murein hydrolase).
GN LYTB OR SP0965.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
RC STRAIN=R6;
RA MEDLINE=99195827; PubMed=10096093;
RA Garcia P., Gonzalez M.P., Garcia E., Lopez R., Garcia J.L.;
RT "lytb, a novel pneumococcal murein hydrolase essential for cell
RT separation.";
RL M01. Microbiol. 31:1275-1281(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TRGR4;
RA MEDLINE=21357209; PubMed=11463916;
RA Tettein H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolony J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feidblum T.V., Anguino S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus

RT pneumonae.";
RL Science 293:498-506(2001).
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
CC CELL SEPARATION.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC unit in high-mannose glycopeptides and glycoproteins containing
CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC residue remains attached to the protein; the rest of the
CC oligosaccharide is released intact.
CC -1- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AJ010312; CAA09078.1; -.
DR EMBL; AE007400; AAK75086.1; -.
DR TRGR; SP0965; -.
DR InterPro; IPR002901; Amidase_4.
DR Pfam; PF01832; Amidase_4; 1.
DR SMART; SM00047; LY22; 1.
KM Signal; Hydrolase; Cell wall; Complete proteome.
FT SIGNAL 1 23
FT CHAIN 24 658
FT FT 336 336 PUTATIVE ENDO-BETA-N-
FT CONFLICT 381 381 ACTYLGLUCOSAMINIDASE.
FT CONFLICT 384 384 T -> M (IN REF. 1).
FT CONFLICT 384 384 E -> A (IN REF. 1).
FT CONFLICT 535 535 E -> K (IN REF. 1).
FT CONFLICT 535 535 L -> P (IN REF. 1).
FT CONFLICT 580 580 F -> S (IN REF. 1).
SQ SEQUENCE 658 AA; 76469 MW; B62515006C9C836 CRC64;

Query Match 24.3%; Score 54.5; DB 1; Length 658;
Best Local Similarity 27.1%; Pred. No. 9.8;
Matches 13; Conservative 9; Mismatches 15; Indels 11; Gaps 1;

OY 3 YKMYRYALKSGRLNP-----MPENCAANNMVFEGTQ 39
:|:|||||:|
DB 213 FENGHYYLYLKGGMANEMWIDKESWFLKFDGKMAKEMVYDSHQ 260

RESULT 14
RN12_YEAST STANDARD; PRT; 850 AA.
ID RN12_YEAST
AC P32843;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE RNA12 protein.
GN RNA12 OR PRP12 OR YMR302C OR YM9952.04C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 28583 / FL100;
RA MEDLINE=92212295; PubMed=1557037;
RA Liang S., Alksne L., Warner J.R., Lacroute F.;
RT "RNA12+, a gene of Saccharomyces cerevisiae involved in pre-tRNA
RT maturation. Characterization of a temperature-sensitive mutant,
RT cloning and sequencing of the gene.";
RL M01. Gen. Genet. 232:304-312(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RA Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS A DISPENSABLE ROLE IN EARLY MATURATION OF

```
CC PRE-RNA. IT MAY PROVIDE A COFACTOR WHICH PLAYS A MINOR AND
CC DISSENSIBLE ROLE IN PRE-RNA MATURATION. THIS PROTEIN MIGHT
CC MIGRATE PROGRESSIVELY ALONG EACH PRE-RNA MOLECULE IN THE EARLY
CC STEPS OF THE RNA MATURATION PATHWAY AND BE RELEASED ONCE THE
CC MODIFICATION IT CATALYSES HAD BEEN COMPLETED.
CC -1- SIMILARITY: TO S.POMBE SPEC83.05.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S92205; AAB21991.1; -
DR EMBL: Z49212; CAA89135.1; -
DR PIR: S20462; S20462.
DR SGD: S0004917; PRP12.
DR InterPro: IPR000504; RNA_rec_mot.
DR Pfam: PF00076; rrm; 1.
DR SMART: SM00360; RRM; 1.
FT VARIANT 502 502
FT N -> Y (IN RNA12-1; DOMINANT TS
FT PHENOTYPE, EXHIBITS ABERRANT PROCESSING
FT OF PRE-RRNA AND UNDERMETHYLATION OF
FT NASCENT RRNA).
SQ SEQUENCE 850 AA; 9668 MW; 3BERD730AD376C7 CRC64;

OY 4 KNYRYALKS-GGRLNAPMPCNGVAENND 31
Db 597 EDLYLNKSKSGENKPESEKETAEKND 625

RESULT 15
CARB_ARCFU STANDARD; PRT: 1076 AA.
ID CARB_ARCFU
AC 028994;
DT 15-JUN-2002 (rel. 41, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
DE phosphate synthetase ammonia chain).
GN CARB OR AF1274.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobi; Archaeoglobales;
OC Archaeoglobaceae; Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Kleink H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Krelavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus S.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Arlrich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RA "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
CC -1- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + L-glutamate + carbamoyl phosphate.
CC -1- COFACTOR: Binds three manganese ions (By similarity).
CC -1- PATHWAY: Arginine biosynthesis.
```

```
CC -1- PATHWAY: Pyrimidine biosynthesis; first step.
CC -1- SUBUNIT: Composed of two chains; the small (or glutamine) chain
CC promotes the hydrolysis of glutamine to ammonia, which is used by
CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
CC similarity).
CC -1- SIMILARITY: BELONGS TO THE CARB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AEO01016; AAB89970.1; -
DR HSSP: P00968; 1A9X.
DR TIGR: AF1274; -
DR InterPro: IPR005483; CPase_L.
DR InterPro: IPR005479; CPase_L_D2.
DR InterPro: IPR005480; CPase_L_D3.
DR InterPro: IPR005481; CPase_L_N.
DR InterPro: IPR004362; MGS_1-like.
DR Pfam: PF00289; CPase_L_chain; 2.
DR Pfam: PF02786; CPase_L_D2; 2.
DR Pfam: PF02787; CPase_L_D3; 1.
DR Pfam: PF02142; MGS; 1.
DR PRINTS: PR00098; CPASE.
DR PROSITE: PS00866; CPASE_1; 2.
DR PROSITE: PS00867; CPASE_2; 1.
KW Arginine biosynthesis; Pyrimidine biosynthesis; Ligase; Repeat;
KW ATP-binding; Manganese; Complete proteome.
FT DOMAIN 1 402 CARBOXYPHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 403 555 OLIGOMERIZATION DOMAIN.
FT DOMAIN 556 939 CARBAMOYL PHOSPHATE SYNTHETIC DOMAIN.
FT DOMAIN 940 1076 ALLOSTERIC DOMAIN.
FT REPEAT 1 555
FT REPEAT 556 1076
FT NP_BIND 153 210
FT NP_BIND 303 353
FT METAL 285 285 ATP (POTENTIAL).
FT METAL 299 299 MANGANESE 1 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 1 AND 2 (BY SIMILARITY).
FT METAL 301 301 MANGANESE 2 (BY SIMILARITY).
FT METAL 830 830 MANGANESE 3 (BY SIMILARITY).
FT METAL 842 842 MANGANESE 3 (BY SIMILARITY).
SQ SEQUENCE 1076 AA; 119469 MW; 3FC275F648E5949D CRC64;

OY 16 RLNAPMPCNGVAEN 29
Db 685 RLNAPMPCNGVAEN 698

Query Match 23.7%; Score 53; DB 1; Length 1076;
Best Local Similarity 64.3%; Pred. No. 28;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

Search completed: December 10, 2002, 10:54:42
Job time : 7.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:49:03 ; Search time 21.8 Seconds
(without alignments)
378.068 Million cell updates/sec

Title: US-09-142-970-4_COPY_1_40

Perfect score: 224

Sequence: 1 LYKNRYALKSGRLNAPMPENGVAENNDVFMGTQE 40

Scoring table: BLOSUM62

Gapop.10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	211	94.2	993	2	Q9S6X3
2	211	94.2	996	2	Q30574
3	211	94.2	997	2	Q9S6X2
4	211	94.2	1552	2	Q9A106
5	199	88.8	996	2	Q57309
6	189	84.4	992	2	Q57035
7	189	84.4	992	2	Q9S6X5
8	189	84.4	992	2	Q9S6X4
9	189	84.4	992	2	Q30573
10	189	84.4	997	2	Q30575
11	189	84.4	1561	2	Q51169
12	189	84.4	1773	16	Q9JVB9
13	185	82.6	1815	16	Q9K0B4
14	101	45.1	1764	2	Q93T34
15	61	27.2	452	16	Q82RU6
16	61	27.2	452	16	Q829H2

17	61	27.2	452	16	Q8X9Z1	Q8X9Z1 escherichia
18	57	25.4	217	17	Q96XX6	Q96XX6 sulfolobus
19	56	25.0	161	11	Q8VEA3	Q8VEA3 mus musculus
20	56	25.0	189	17	Q26218	Q26218 methanobact
21	56	25.0	1367	5	Q9TYU0	Q9TYU0 caenorhabd
22	55.5	24.8	435	5	Q9NFT4	Q9NFT4 trichomonas
23	55.5	24.8	948	5	Q9V9Q0	Q9V9Q0 drosophila
24	55	24.6	350	2	Q43983	Q43983 acinetobact
25	55	24.6	573	2	Q32433	Q32433 acinetobact
26	55	24.6	574	16	Q8XEC2	Q8XEC2 escherichia
27	54.5	24.3	455	5	Q24464	Q24464 drosophila
28	54.5	24.3	459	5	Q9NGW6	Q9NGW6 drosophila
29	54.5	24.3	459	5	Q9VW46	Q9VW46 drosophila
30	54.5	24.3	463	5	Q9NBK5	Q9NBK5 drosophila
31	54.5	24.3	614	2	Q9AHT8	Q9AHT8 streptococc
32	54.5	24.3	766	5	Q18673	Q18673 caenorhabd
33	54	24.1	283	9	Q9AF60	Q9AF60 streptococc
34	54	24.1	473	16	Q8RD72	Q8RD72 fusobacteri
35	54	24.1	630	17	Q8ZTA8	Q8ZTA8 pyrobaculum
36	54	24.1	1363	12	Q8V436	Q8V436 bovine coro
37	54	24.1	2146	13	Q9PMU3	Q9PMU3 gallus gall
38	54	24.1	2157	13	Q9PMU4	Q9PMU4 gallus gall
39	54	24.1	2171	13	Q9PMU6	Q9PMU6 gallus gall
40	54	24.1	2182	13	Q9PMU5	Q9PMU5 gallus gall
41	54	24.1	2321	13	Q9PMU4	Q9PMU4 gallus gall
42	54	24.1	2332	13	Q9PMU5	Q9PMU5 gallus gall
43	54	24.1	2346	13	Q9PMU7	Q9PMU7 gallus gall
44	54	24.1	2357	13	Q9PMU6	Q9PMU6 gallus gall
45	53.5	23.9	660	2	Q57443	Q57443 haemophilus

ALIGNMENTS

RESULT 1
Q9S6X3 ID Q9S6X3 PRELIMINARY; PRT; 993 AA.
AC Q9S6X3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Igal protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24400;
RX MEDLINE=98010345; PubMed=9350862;
RA MORELLI G., Malorny B., Muller K., Selter A., Wang J.F., del Valle J., Achtmann M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30 years of epidemic spread."
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012208; AAC45791.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASRPRTASE.
DR PROSITE: PS00462; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 993
SQ SEQUENCE 993 AA; 109441 MW; 109FAA2EF8AC3C6 CRC64;

Query Match 94.2%; Score 211; DB 2; Length 993;
Best Local Similarity 92.5%; Pred. No. 7.1e-20;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGRLNAPMPENGVAENNNDWFMGTYOE 40
|||||
DB 558 LYKKNRYVALKSGGSYNAPMPENGVTENNNDWFMGTYOE 597

RESULT 2

ID 030574 PRELIMINARY; PRT; 996 AA.
AC 030574;
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 1gal protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24024;
RX MEDLINE=98010345; PubMed=9350862;
RA MORELLI G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL, AF012206; AAC45789.2; -
DR MEROPS, S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Protease.
FT NON_TER 1 1
FT NON_TER 996 996
SQ SEQUENCE 996 AA; 109688 MW; B3ABDEF54C337D9 CRC64;

Query Match 94.2%; Score 211; DB 2; Length 996;
Best Local Similarity 92.5%; Pred. No. 7.2e-20;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGRLNAPMPENGVAENNNDWFMGTYOE 40
|||||
DB 557 LYKKNRYVALKSGGSYNAPMPENGVTENNNDWFMGTYOE 596

RESULT 3
ID 0956X2 PRELIMINARY; PRT; 997 AA.
AC 0956X2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 1gal protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24081;
RX MEDLINE=98010345; PubMed=9350862;
RA MORELLI G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL, AF012209; AAC45792.2; -
DR MEROPS, S06.001; -
DR InterPro: IPR002195; Dihydroorotase.

DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Protease.
FT NON_TER 1 1
FT NON_TER 997 997
SQ SEQUENCE 997 AA; 109812 MW; 06F2E361E7E202E0 CRC64;
Query Match 94.2%; Score 211; DB 2; Length 997;
Best Local Similarity 92.5%; Pred. No. 7.2e-20;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

RESULT 4

ID 09A1U6 PRELIMINARY; PRT; 1552 AA.
AC 09A1U6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 1gal protease.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RA Vitoyski S., Sayers J.B.;
RT "Degenerate specificity of Neisseria meningitidis 1gal protease."
RL Submitted (FEB-2000) to the EMBL/Genbank/DBD databases.
DR EMBL, AF285032; AAK15073.1; -
DR MEROPS, S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Protease.
SQ SEQUENCE 1552 AA; 171205 MW; E49D6657B120CEFF CRC64;

Query Match 94.2%; Score 211; DB 2; Length 1552;
Best Local Similarity 92.5%; Pred. No. 1.2e-19;
Matches 37; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 LYKKNRYVALKSGRLNAPMPENGVAENNNDWFMGTYOE 40
|||||
DB 574 LYKKNRYVALKSGGSYNAPMPENGVTENNNDWFMGTYOE 613

RESULT 5
ID 057309 PRELIMINARY; PRT; 996 AA.
AC 057309;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 1gal protease precursor (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23524;

RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achtmann M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread.";
 RL Mol. Microbiol. 25:1047-1064(1997).
 RN (2)
 RP SEQUENCE OF 37-532 FROM N.A.
 RC STRAIN-ETH2;
 RX MEDLINE-95302961; PubMed-7783620;
 RA Lomholt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding Iga1 protease
 RT in *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Haemophilus*
 RT influenzae.";
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL; X82469; CAAS57852.1; -;
 DR EMBL; X82468; CAAS57851.1; -;
 DR MEROPS; S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertactin-sup.
 DR Pfam; PF03395; IGA1; 1.
 DR Pfam; PF03212; Pertactin; 1.
 DR PRINTS; PR00921; IGASERPTASE.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KW Protease.
 FT NON_TER
 FT SEQUENCE 996 AA; 109717 MW; AB39F7B98A41F986 CRC64;
 SO

Query Match 88.8%; Score 199; DB 2; Length 996;
 Best Local Similarity 87.5%; Pred. No. 3,1e-18;
 Matches 35; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGRLNAMPENGVAENNDVFMGYTOE 40
 ||||||||||||| :||||||| ||||||||| ||
 DB 557 LYKKNRYVALKSGSVNAMPENGQTEENNNDVFMGYKOE 596

RESULT 6
 O57035 PRELIMINARY; PRT; 992 AA.
 AC O57035;
 DT 01-NOV-1996 (TREMblrel. 01, Created)
 DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Iga1 protease precursor (Fragment).
 GN IGA.
 OS *Neisseria meningitidis*.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_TaxID=487;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-B40;
 RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F.,
 RA del Valle J., Achtmann M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread.";
 RL Mol. Microbiol. 25:1047-1064(1997).
 RN (2)
 RP SEQUENCE OF 37-532 FROM N.A.
 RC STRAIN-HF48;
 RX MEDLINE-95302961; PubMed-7783620;
 RA Lomholt H., Poulsen K., Mogens K.;
 RT "Comparative characterization of the iga gene encoding Iga1 protease
 RT in *Neisseria meningitidis*, *Neisseria gonorrhoeae* and *Haemophilus*
 RT influenzae.";
 RL Mol. Microbiol. 15:495-506(1995).
 DR EMBL; AF012211; AAC45794.2; -;
 DR EMBL; X82480; CAAS7863.1; -;
 DR EMBL; X82475; CAAS7858.1; -;

DR MEROPS; S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertactin-sup.
 DR Pfam; PF03395; IGA1; 1.
 DR Pfam; PF03212; Pertactin; 1.
 DR PRINTS; PR00921; IGASERPTASE.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KW Protease.
 FT NON_TER
 FT SEQUENCE 992 AA; 109228 MW; 3677DDEACE6D9F69 CRC64;
 SO

Query Match 84.4%; Score 189; DB 2; Length 992;
 Best Local Similarity 82.5%; Pred. No. 7e-17;
 Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGRLNAMPENGVAENNDVFMGYTOE 40
 ||||||||||||| :||||||| ||||||||| ||
 DB 557 LYKKNRYVALKSGSVNAMPENGQTEENNNDVFMGSTOE 596

RESULT 7
 O9S6X5 PRELIMINARY; PRT; 992 AA.
 AC O9S6X5;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE Iga1 protease (Fragment).
 GN IGA.
 OS *Neisseria meningitidis*.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; *Neisseria*.
 OX NCBI_TaxID=487;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-42491;
 RX MEDLINE-98010345; PubMed-9350862;
 RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
 RA Achtmann M.;
 RT "Clonal descent and microevolution of *Neisseria meningitidis* during 30
 RT years of epidemic spread.";
 RL Mol. Microbiol. 25:1047-1064(1997).
 DR EMBL; AF012203; AAC45786.1; -;
 DR MEROPS; S06.001; -;
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; Iga_S6.
 DR InterPro: IPR004899; Pertactin-sup.
 DR Pfam; PF03395; IGA1; 1.
 DR Pfam; PF03212; Pertactin; 1.
 DR PRINTS; PR00921; IGASERPTASE.
 DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 KW Protease.
 FT NON_TER
 FT SEQUENCE 992 AA; 109239 MW; 306FF451B7213D49 CRC64;
 SO

Query Match 84.4%; Score 189; DB 2; Length 992;
 Best Local Similarity 82.5%; Pred. No. 7e-17;
 Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGRLNAMPENGVAENNDVFMGYTOE 40
 ||||||||||||| :||||||| ||||||||| ||
 DB 557 LYKKNRYVALKSGSVNAMPENGQTEENNNDVFMGSTOE 596

RESULT 8
 O9S6X4 PRELIMINARY; PRT; 992 AA.
 AC O9S6X4;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)

DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread.";
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012204; AAC45787.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER
FT NON_TER
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CE3 CRC64;

Query Match 84.4%; Score 189; DB 2; Length 992;
Best Local Similarity 82.5%; Pred. No. 7e-17;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYYKNRYALKSGRLNAMPENGVAENNDVFMGYTQE 40
DB 557 LYYKNRYALKSGGSVNAPEMNGOTENNNDWILMGSTQE 596

RESULT 9
ID 030573 PRELIMINARY; PRT; 992 AA.
AC 030573;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23910;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread.";
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012205; AAC45788.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER
FT NON_TER
SQ SEQUENCE 992 AA; 109255 MW; EF7CA563BC7D7ECB CRC64;

Query Match 84.4%; Score 189; DB 2; Length 992;

Best Local Similarity 82.5%; Pred. No. 7e-17;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYYKNRYALKSGRLNAMPENGVAENNDVFMGYTQE 40
DB 557 LYYKNRYALKSGGSVNAPEMNGOTENNNDWILMGSTQE 596

RESULT 10
ID 030575 PRELIMINARY; PRT; 997 AA.
AC 030575;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24099;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
Achtman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread.";
RL MOL. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012210; AAC45793.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER
FT NON_TER
SQ SEQUENCE 997 AA; 110152 MW; 4DE40DE594E5E28E CRC64;

Query Match 84.4%; Score 189; DB 2; Length 997;
Best Local Similarity 82.5%; Pred. No. 7e-17;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYYKNRYALKSGRLNAMPENGVAENNDVFMGYTQE 40
DB 558 LYYKNRYALKSGGSVNAPEMNGOTENNNDWILMGSTQE 597

RESULT 11
ID 051169 PRELIMINARY; PRT; 1561 AA.
AC 051169;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease.
GN IGA.
OS Neisseria meningitidis.
OC Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF13;
RX MEDLINE=95302961; PubMed=7783620;
RA Lombolt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
influenzae.";
RL MOL. Microbiol. 15:495-506(1995).

DR EMBL: X82474; CAA57857.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Protease.
KM
SQ SEQUENCE 1561 AA; 171849 MW; 1C96E291A00017D5 CRC64;

Query Match
Best Local Similarity 84.4%; Score 189; DB 2; Length 1561;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAPMPENGVAENNDWFMGYTOE 40
DB 584 LYKKNRYVALKSGGSVNAAMPENGOTENNNDWILMGSTOE 623

RESULT 12
O9JVB9 PRELIMINARY; PRT: 1773 AA.
AC O9JVB9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Iga1 protease (EC 3.4.21.7).
GN IGA OR NMA0905.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SERO TYPE 4A;
RX MEDLINE=2022356; PubMed=10761919;
RA Parhill J., Achman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S., Jagsels K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.";
RT Nature 404:502-506(2000).
RL EMBL: AL162754; CAB84182.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
DR Protease; Hydrolase; Complete proteome.
KM
SQ SEQUENCE 1773 AA; 196350 MW; CAC19E7313D76CE1 CRC64;

Query Match
Best Local Similarity 84.4%; Score 189; DB 16; Length 1773;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAPMPENGVAENNDWFMGYTOE 40
DB 573 LYKKNRYVALKSGGSVNAAMPENGOTENNNDWILMGSTOE 612

RESULT 13
O9KOB4 PRELIMINARY; PRT: 1815 AA.
AC O9KOB4;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Iga-specific serine endopeptidase.
GN NMB0700.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eissen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parksey D.S., Blair E., Citron H., Clark E.B., Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Masignani V., Plaza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappelli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RT Science 287:1809-1815(2000).
DR EMBL: AE002424; AAF4117.1; -
DR MEROPS: S06.001; -
DR TIGR: NMB0700; -
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Complete proteome.
SQ SEQUENCE 1815 AA; 201077 MW; 2259D4D71762C57F CRC64;

Query Match
Best Local Similarity 82.6%; Score 185; DB 16; Length 1815;
Matches 32; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAPMPENGVAENNDWFMGYTOE 40
DB 573 LYKKNRYVALKSGGSVNAAMPENGOTENNNDWILMGSTOE 612

RESULT 14
O93T34 PRELIMINARY; PRT: 1764 AA.
AC O93T34;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE Iga1 protease type 2.
GN IGA1.
OS Haemophilus aegyptius.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=725;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F3031;
RA McGillivray G., Actis L.A.;
RT "Iga protease from H. aegyptius F3031.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF369907; AAK56925.1; -
DR MEROPS: S06.001; -
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR Protease.
KM
SQ SEQUENCE 1764 AA; 193863 MW; EC583CDBE81DBEC CRC64;

Query Match
Best Local Similarity 45.1%; Score 101; DB 2; Length 1764;
Matches 32; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Comugen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:14:59 : Search time 28.2 seconds
(without alignments)
189.008 Million cell updates/sec

Title: US-09-142-970-5_COPY_1_40

Perfect score: 224
Sequence: 1 LYKNRYRYALKSGRLNAPMPENGVAENNDIMFGYQOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq.101002.*
1: /SID2/gcgdata/geneseq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	100.0	104	AAW61606	Neisseria IgA1 pro
2	223	99.6	104	AAW61605	Neisseria IgA1 pro
3	210	93.8	104	AAW61603	Neisseria IgA1 pro
4	198	88.4	104	AAW61604	Neisseria IgA1 pro
5	190	84.8	105	AAW65656	105-mer peptide us
6	190	84.8	105	AAW61602	Neisseria IgA1 pro
7	93	41.5	1541	AAW07304	IgA1 protease, Ha
8	62	27.7	452	AAU34434	E. coli cellular p
9	62	27.7	452	AAU38481	Salmonella typhi c
10	62	27.7	452	AAW98402	Escherichia coli p

11	62	27.7	834	AAW34542	Porphyromonas ging
12	62	27.7	907	AAW34408	Porphyromonas ging
13	60.5	27.0	971	AAW95687	Cosmid CHR155 enc
14	57	25.4	3084	AAW50891	Mouse laminin A ch
15	57	25.4	3084	AAE11215	Mouse laminin-1 al
16	55.5	24.8	948	ABW58055	Drosophila melanog
17	55	24.6	569	AAW61214	Streptococcus pneu
18	55	24.6	569	ABP54633	S. pneumoniae SP08
19	55	24.6	591	AAW55099	Streptococcus pneu
20	55	24.6	591	ABP54593	S. pneumoniae SP04
21	55	24.6	678	AAW81667	Streptococcus pneu
22	55	24.6	1216	ABG09754	Novel human diagno
23	55	24.6	2901	ABG09763	Human laminin A ch
24	55	24.6	3075	AAW50892	Human OREF protein
25	53.5	23.9	194	ABP07412	Truncated transfer
26	53.5	23.9	265	AAW43003	Truncated transfer
27	53.5	23.9	265	AAW51810	H. influenzae type
28	53.5	23.9	310	AAW43004	Truncated transfer
29	53.5	23.9	310	AAW51809	H. influenzae type
30	53.5	23.9	365	AAW43005	Truncated transfer
31	53.5	23.9	365	AAW51808	H. influenzae type
32	53.5	23.9	404	AAW43006	Truncated transfer
33	53.5	23.9	404	AAW51807	H. influenzae type
34	53.5	23.9	411	AAW43007	Truncated transfer
35	53.5	23.9	411	AAW51806	H. influenzae type
36	53.5	23.9	417	AAW43008	Truncated transfer
37	53.5	23.9	417	AAW51805	H. influenzae type
38	53.5	23.9	425	AAW43009	Truncated transfer
39	53.5	23.9	430	AAW51804	H. influenzae type
40	53.5	23.9	456	AAW8365	Drosophila Ndr ser
41	53.5	23.9	459	ABW1664	Drosophila melanog
42	53.5	23.9	463	AAW43010	Truncated transfer
43	53.5	23.9	463	AAW51803	H. influenzae type
44	53.5	23.9	523	AAW43011	Truncated transfer
45	53.5	23.9	529	AAW43012	Truncated transfer

ALIGNMENTS

RESULT 1
AAW61606
ID AAW61606 standard; peptide: 104 AA.
XX
AC AAW61606;
XX
DT 27-OCT-1998 (first entry)
XX
DE Neisseria IgA1 protease fragment 5.
XX
KW Immunoglobulin protease; carrier; paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
OS Neisseria sp.
XX
PN MO9831791-A1.
XX
PD 23-JUL-1998.
XX
PF 20-JAN-1998; 98WO-EP00294.
XX
PR 21-JAN-1997; 97EP-0100883.
XX
PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
PI Achmann M, Moreau M;
XX
DR WPI, 1998-414092/35.
XX
PT New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and

PT Haemophilus
 XX
 PS Claim 2: Fig 4: 32pp; English.
 CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.
 XX
 SQ Sequence 104 AA;
 Query Match 100.0%; Score 224; DB 19; Length 104;
 Best Local Similarity 100.0%; Pred. No. 5e-23;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYYKNRYVALKSGRLNAPMPENGVAENNDWIFMGYTQ 40
 DB 1 LYYKNRYVALKSGRLNAPMPENGVAENNDWIFMGYTQ 40
 RESULT 2
 AAW61605 standard; peptide; 104 AA.
 XX AAW61605;
 AC
 XX 27-OCT-1998 (first entry)
 DT
 XX
 DE Neisseria IgA1 protease fragment 4.
 XX
 KW Immunoglobulin protease; carrier; paediatric; vaccine;
 KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
 XX
 OS Neisseria sp.
 XX
 PN WO9831791-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 20-JAN-1998; 98WO-EP00294.
 XX
 PR 21-JAN-1997; 97EP-0100883.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
 PI Achtmann M, Moreau M;
 XX
 DR WPI: 1998-414092/35.
 XX
 PT New peptide from Neisseria immunoglobulin protease - useful as
 PT immunogenic carrier, e.g. particularly for polysaccharide(s),
 PT forming conjugates used in vaccines against Neisseria and
 PT Haemophilus
 XX
 PS Claim 2: Fig 1/4: 32pp; English.
 XX
 CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.
 XX
 SQ Sequence 104 AA;
 Query Match 99.6%; Score 223; DB 19; Length 104;

Best Local Similarity 97.5%; Pred. No. 6.9e-23;
 Matches 39; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LYYKNRYVALKSGRLNAPMPENGVAENNDWIFMGYTQ 40
 DB 1 LYYKNRYVALKSGRLNAPMPENGVAENNDWIFMGYTQ 40
 RESULT 3
 AAW61603 standard; peptide; 104 AA.
 XX AAW61603;
 AC
 XX 27-OCT-1998 (first entry)
 DT
 XX
 DE Neisseria IgA1 protease fragment 2.
 XX
 KW Immunoglobulin protease; carrier; paediatric; vaccine;
 KM epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
 XX
 OS Neisseria sp.
 XX
 PN WO9831791-A1.
 XX
 PD 23-JUL-1998.
 XX
 PF 20-JAN-1998; 98WO-EP00294.
 XX
 PR 21-JAN-1997; 97EP-0100883.
 XX
 PA (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
 PI Achtmann M, Moreau M;
 XX
 DR WPI: 1998-414092/35.
 XX
 PT New peptide from Neisseria immunoglobulin protease - useful as
 PT immunogenic carrier, e.g. particularly for polysaccharide(s),
 PT forming conjugates used in vaccines against Neisseria and
 PT Haemophilus
 XX
 PS Claim 2: Fig 2: 32pp; English.
 XX
 CC The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
 CC as carriers for a conjugate, particularly in combination with a
 CC polysaccharide. They can be used in paediatric or other vaccines,
 CC particularly for prevention of epidemic bacterial infections, especially
 CC those caused by Neisseria or Haemophilus. The protease fragment is a
 CC highly immunogenic carrier that elicits a T-cell response, resulting in
 CC a long-lasting memory and high antibody titre, and possibly making
 CC possible vaccination without adjuvant.
 XX
 SQ Sequence 104 AA;
 Query Match 93.8%; Score 210; DB 19; Length 104;
 Best Local Similarity 90.0%; Pred. No. 4.2e-21;
 Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 QY 1 LYYKNRYVALKSGRLNAPMPENGVAENNDWIFMGYTQ 40
 DB 1 LYYKNRYVALKSGSVNAPMPENGTENNNDWIFMGYTQ 40
 RESULT 4
 AAW61604 standard; peptide; 104 AA.
 XX AAW61604;
 AC
 XX 27-OCT-1998 (first entry)
 DT
 XX

DE Neisseria IgA1 protease fragment 3.
XX
XX Immunoglobulin protease; carrier; paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
XX
XX MO9831791-A1.
XX
XX 23-JUL-1998.
XX
XX 20-JAN-1998; 98WO-EP00294.
XX
XX 21-JAN-1997; 97EP-0100883.
XX
XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
XX
XX WPI; 1998-414092/35.
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 2; Fig 1; 32pp; English.
XX
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 104 AA:
SQ

Query Match 88.4%; Score 198; DB 19; Length 104;
Best Local Similarity 85.0%; Pred. No. 1.9e-19;
Matches 34; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGGRNAPMPENGVAENNDWIFMGYTOE 40
Db 1 LYYKNRYVALKSGSVNAPMPENGOTENNNDWVFMGYKOE 40

RESULT 5
AAW65566
ID AAW65566 standard; peptide; 105 AA.
XX
XX AAW65566;
XX
XX 15-OCT-1998 (first entry)
XX
XX 105-mer peptide used in polysaccharide-peptide conjugate.
DE
XX
KW Polysaccharide-peptide conjugate; vaccine; linker moiety; adjuvant;
KW Immune response.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
FT Modified-site 1 /note- "N-terminal acetyl"
FT
XX
XX MO9831393-A2.
XX
XX 23-JUL-1998.
XX
XX 21-JAN-1998; 98WO-EP00654.

XX
XX 21-JAN-1997; 97EP-0100884.
XX
XX (INMR) PASTEUR MERIEUX SERUMS & VACCINS SA.
PA
XX
XX Mistretta N, Moreau M;
XX
XX WPI; 1998-413820/35.
XX
XX Polysaccharide-peptide conjugates, useful as, e.g. vaccines -
PT comprise peptide moiety with at least six amino acid residues,
PT polysaccharide chain with at least four repeat units, and linker
PT moiety
XX
XX Example 1; Page 14; 28pp; English.
XX
XX The invention relates to: (A) polysaccharide-peptide conjugate (in which
CC the polysaccharide is immunogenic), comprising: (a) a peptide moiety
CC which has at least 6 amino acid residues, at least 1 of which is a
CC cysteine residue; (b) a polysaccharide chain comprising at least 4 repeat
CC units, and (c) a linker moiety bound to the thiol group of the cysteine.
CC The linker is also bound to: (i) native amino, hydroxyl or carboxyl
CC groups of the polysaccharide chain; (ii) amino groups created by
CC hydrolysis of native N-acyl groups of the polysaccharide chain, or (iii)
CC functional groups introduced on the polysaccharide chain upon
CC derivatisation with a spacer moiety bound to native amino, hydroxyl or
CC carboxyl groups of the polysaccharide chain, and (b) conjugating a
CC peptide as in (Aa) to a polysaccharide chain comprising at least 4 repeat
CC units, comprising: (a) coupling the peptide to a linker through the thiol
CC group of the cysteine residue, and (b) coupling the linker to the
CC polysaccharide chain through one of groups (i), (ii) or (iii) as
CC described in (A). The conjugates are especially useful as vaccines to
CC elicit a protective long term immune response against a pathogenic
CC microorganism from which the immunogenic polysaccharide chain is derived.
CC Known polysaccharide-peptide conjugates are less immunogenic than their
CC polysaccharide-protein counterparts and require adjuvantation. The new
CC conjugates have good immunogenicity, largely because the peptides are of
CC sufficient size. The present sequence represents a peptide used in
CC a polysaccharide-peptide conjugate.
XX
XX Sequence 105 AA:
SQ

Query Match 84.8%; Score 190; DB 19; Length 105;
Best Local Similarity 85.0%; Pred. No. 2.3e-18;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYYKNRYVALKSGGRNAPMPENGVAENNDWIFMGYTOE 40
Db 2 LYYKNRYVALKSGSVNAPMPENGOTENNNDWILMGSTOE 41

RESULT 6
AAW61602
ID AAW61602 standard; peptide; 105 AA.
XX
XX AAW61602;
XX
XX 27-OCT-1998 (first entry)
XX
XX Neisseria IgA1 protease fragment 1.
DE
XX
KW Immunoglobulin protease; carrier; paediatric; vaccine;
KW epidemic bacterial infection; Neisseria; Haemophilus; T-cell.
XX
XX Neisseria sp.
XX
XX MO9831791-A1.
XX
XX 23-JUL-1998.
XX
XX 20-JAN-1998; 98WO-EP00294.
XX
XX 21-JAN-1997; 97EP-0100883.

```

XX (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PA (INMR ) PASTEUR MERIEUX SERUMS & VACCINS SA.
XX
XX Achtmann M, Moreau M;
XX WPI; 1998-414092/35.
XX
XX New peptide from Neisseria immunoglobulin protease - useful as
PT immunogenic carrier, e.g. particularly for polysaccharide(s),
PT forming conjugates used in vaccines against Neisseria and
PT Haemophilus
XX
XX Claim 6; Page 10; 32pp; English.
XX
XX The Neisseria immunoglobulin protease fragments AAW61602-W61606 are used
CC as carriers for a conjugate, particularly in combination with a
CC polysaccharide. They can be used in paediatric or other vaccines,
CC particularly for prevention of epidemic bacterial infections, especially
CC those caused by Neisseria or Haemophilus. The protease fragment is a
CC highly immunogenic carrier that elicits a T-cell response, resulting in
CC a long-lasting memory and high antibody titre, and possibly making
CC possible vaccination without adjuvant.
XX
XX Sequence 105 AA;
SQ
Query Match 84.8%; Score 190; DB 19; Length 105;
Best Local Similarity 85.0%; Pred. No. 2.3e-18;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWIMGYTQ 40
Db 2 LYYKNRYALKSGSVNAPMPENGTEENNDWIMGSTQ 41
RESULT 7
AAU07304
ID AAU07304 standard; protein; 1541 AA.
XX
XX AAU07304;
XX
XX 31-JAN-1991 (first entry)
XX
XX TgaI protease.
XX
XX IgA1; vaccine; meningitis; gonorrhoea; allergies.
XX
XX Haemophilus influenzae.
XX
XX WO9011367-A.
XX
XX 04-OCT-1990.
XX
XX 16-MAR-1990; 90WO-DK00073.
XX
XX 17-MAR-1989; 89DK-0001308.
XX
XX (KILI/) KILIAN M.
XX
XX Kilian M, Poulsen K;
XX
XX WPI; 1990-320267/42.
XX
XX N-PSDB; AAQ06164.
XX
XX Immunoglobulin A1 protease prodn. - by cloning from
PT microorganisms for immunisation against immunoglobulin A1
PT protease producing bacteria
XX
XX Disclosure; fig 3; 44pp; English.
XX
XX This immunoglobulin (Ig)A1 protease is produced by recombinant DNA
CC methods. It is useful in a vaccine for e.g. meningococcal meningitis,
CC gonorrhoea or allergic diseases. It specifically cleaves the heavy

```

```

CC chain of human IgA1 in the hinge region.
XX
XX Sequence 1541 AA;
SQ
Query Match 41.5%; Score 93; DB 11; Length 1541;
Best Local Similarity 37.5%; Pred. No. 0.0011;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;
QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWIMGYTQ 40
Db 595 LNEENTYALKRGASTRSELPKSGSENEWLYMGKTS 634
RESULT 8
AAU34434
ID AAU34434 standard; Protein; 452 AA.
XX
XX AAU34434;
XX
XX 14-FEB-2002 (first entry)
XX
XX E. coli cellular proliferation protein #15.
XX
XX Antisense; prokaryotic cellular proliferation protein;
XX antibiotic; antibacterial; drug design.
XX
XX Escherichia coli.
XX
XX WO200170955-A2.
XX
XX 27-SEP-2001.
XX
XX 21-MAR-2001; 2001WO-US09180.
XX
XX 21-MAR-2000; 2000US-191078P.
XX
XX 23-MAY-2000; 2000US-206848P.
XX
XX 26-MAY-2000; 2000US-207727P.
XX
XX 23-OCT-2000; 2000US-242578P.
XX
XX 27-NOV-2000; 2000US-253625P.
XX
XX 22-DEC-2000; 2000US-257931P.
XX
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
XX Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
XX
XX N-PSDB; AAS52293.
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX Example 3; Seq ID No 10027; 51pp; English.
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at

```


Best Local Similarity 42.3%; Pred. No. 4.6;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;

OY 12 KSGRLNAPMENGVE---AENNDWI 33
ID AAY34542 standard; Protein: 834 AA.
DB 200 KAKEIFSGLEPENGIAINADNDWL 225

RESULT 11

AY34542
ID AAY34542 standard; Protein: 834 AA.
AC AAY34542;
DT 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG71.
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KM vaccine; antigenic.
XX Porphyromonas gingivalis.
OS

XX WO9929870-A1.
XX 17-JUN-1999.
XX 10-DEC-1998; 98WO-AU01023.
XX 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-0001182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002264.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX Agius CT, Barr IG, Hocking DM, Margets MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX WPI: 1999-385613/32.
XX N-PSDB; AAX91760.
DR Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX

PS Claim 1; Page 535-536; 588pp; English.
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX

SO Sequence 834 AA;
Query Match 27.7%; Score 62; DB 20; Length 834;
Best Local Similarity 34.6%; Pred. No. 9.4;
Matches 18; Conservative 7; Mismatches 11; Indels 16; Gaps 4;

OY 3 YKNRY---YALSG-----GRNAPMENGVAE---NNDWI---FMGTT 38
ID AAY95687 standard; Protein: 971 AA.
AC AAY95687;
XX 503 YAGYNYRRKDOYSLDTGLRVSSRLKALFPEENAADFHSNFDWVPLTLGYT 554

RESULT 12

AY34408
ID AAY34408 standard; Protein: 907 AA.
DB 576 YAGYNYRRKDOYSLDTGLRVSSRLKALFPEENAADFHSNFDWVPLTLGYT 627

XX AAY34408;
XX 25-AUG-1999 (first entry)
DE Porphyromonas gingivalis protein PG71.
KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;
KM vaccine; antigenic.
XX Porphyromonas gingivalis.
OS

XX WO9929870-A1.
XX 17-JUN-1999.
XX 10-DEC-1998; 98WO-AU01023.
XX 04-AUG-1998; 98AU-0005028.
XX 10-DEC-1997; 97AU-0000839.
XX 31-DEC-1997; 97AU-0001182.
XX 30-JAN-1998; 98AU-0001546.
XX 10-MAR-1998; 98AU-0002264.
XX 09-APR-1998; 98AU-0002911.
XX 23-APR-1998; 98AU-0003128.
XX 05-MAY-1998; 98AU-0003338.
XX 22-MAY-1998; 98AU-0003654.
XX 29-JUL-1998; 98AU-0004917.
XX (CSLC-) CSL LTD.
XX Agius CT, Barr IG, Hocking DM, Margets MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;
XX WPI: 1999-385613/32.
XX N-PSDB; AAX91626.
DR Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX

PS Claim 1; Page 380-382; 588pp; English.
XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX

SO Sequence 907 AA;
Query Match 27.7%; Score 62; DB 20; Length 907;
Best Local Similarity 34.6%; Pred. No. 10;
Matches 18; Conservative 7; Mismatches 11; Indels 16; Gaps 4;

OY 3 YKNRY---YALSG-----GRNAPMENGVAE---NNDWI---FMGTT 38
ID AAY95687 standard; Protein: 971 AA.
AC AAY95687;
XX 576 YAGYNYRRKDOYSLDTGLRVSSRLKALFPEENAADFHSNFDWVPLTLGYT 627

RESULT 13

AY95687
ID AAY95687 standard; Protein: 971 AA.
AC AAY95687;
XX

```

DT 25-OCT-2000 (first entry)
XX Cosmid CHRIM5 encoded protein P2-0f.
DE Cosmid CHRIM5; nematocide; nematode; biological control agent;
XX transgenic plant; helminthiasis; P2-0f.
XX Xenorhabdus bovienii.
OS MO2000042855-A1.
XX
XX 27-JUL-2000.
XX
XX 24-JAN-2000; 2000WO-GB00219.
XX
XX 22-JAN-1999; 99GB-0001499.
XX
XX (HORT-) HORTICULTURE RES INT.
XX
XX Morgan JAW, Jarrett P, Ellis D, Ousley MA;
XX
XX MPI: 2000-499157/44.
XX
XX N-PSDB; AAA50029.
XX
XX Novel composition used to control parasitic nematodes, especially in
XX plants such as maize, cotton, soya, and rice, comprises a bacterium
XX which is a symbiont of an entomopathogenic nematode -
XX
XX Example 6; Page 38-39; 74pp; English.
XX
XX The present sequence is that of protein P2-0f encoded by an open
XX reading frame identified in cosmid CHRIM5 (see AAA50029). CHRIM5 was
XX obtained by ligating Xenorhabdus bovienii strain I73 (NCIMB 40986)
XX Sau3A-digested DNA fragments into the BamHI site of the Stratagene
XX cosmid vector Supercos1, packaging into Escherichia coli XL Blue 1,
XX and screening for nematocidal activity against Caenorhabditis elegans.
XX Analysis of the DNA indicated a number of open reading frames for
XX which the corresponding protein sequences were determined (see
XX CC AAY95688-Y95735). Nematodes can be controlled through the use of
XX CC bacteria associated symbiotically with an entomopathogenic nematode.
XX Such bacteria include Xenorhabdus and Photorhabdus spp. such as X.
XX bovienii strain I73. The symbiont bacteria, an engineered
XX bacterium, or a nematocidal protein obtained from such bacteria,
XX can be used to control helminthiasis in a human or domesticated
XX CC animal or for the control of plant pathogen nematodes. Also
XX CC claimed are vectors for expressing nematocidal proteins in host
XX cells, and transgenic plants.
XX
XX Sequence 971 AA:
SQ
Query Match 27.0%; Score 60.5; DB 21; Length 971;
Best Local Similarity 26.2%; Pred. No. 18;
Matches 16; Conservative 7; Mismatches 17; Indels 21; Gaps 1;
QY 1 LYYKRYRYALKSGRLNAPM-----PENGVAENNDMIFMGYTQ 39
DB 654 LYYGYRYYPWAGSWLSADPACTIGLNLXRVNRNPNATLDKNGLAPGNRVFFPEIH 713
QY 40 E 40
DB 714 E 714
RESULT 14
AAW50891
ID AAW50891 standard; Protein; 3084 AA.
AC AAW50891;
XX
XX 07-DEC-1998 (first entry)
XX Mouse laminin A chain.
XX

```

```

KW Laminin; mouse; beta-amyloid; amyloidosis; Alzheimer's disease;
KW Down's syndrome; hereditary cerebral haemorrhage; inflammation;
KW malignancy; Familial Mediterranean Fever; multiple myeloma;
KW type II diabetes; prion disease; Creutzfeldt-Jacob disease; CJD;
KW Gertsman-Straussler syndrome; kuru; scrapie; haemodialysis;
KW carpal tunnel syndrome; senile cardiac amyloid polynuropathy;
KW Familial Amyloidotic Polynuropathy; thyroid carcinoma; diagnosis;
KW therapy.
XX
XX Mus sp.
XX
XX OS
XX
XX Key Location/Qualifiers
XX FH 2746..2922
XX FT /note="Fourth globular domain repeat (Claim 13)"
XX FT 2690..2700
XX FT /note="beta-amyloid protein binding region
XX FT (Claim 12)"
XX
XX MO9815179-A1.
XX
XX 16-APR-1998.
XX
XX 08-OCT-1997; 97WO-US18145.
XX
XX 08-OCT-1996; 96US-0027981.
XX
XX (UNITV ) UNITV WASHINGTON.
XX
XX Castillo G, Snow AD;
XX
XX MPI: 1998-240534/21.
XX
XX Use of laminin and fragments - for developing products for use in
XX the diagnosis and treatment of amyloid disease, e.g. Alzheimer's
XX disease or CJD
XX
XX Claim 15; Page 74-79; 132pp; English.
XX
XX This is the amino acid sequence of the mouse laminin A chain. The
XX primary object of the invention is to use laminin, laminin-derived
XX protein fragments and/or laminin-derived polypeptides as potent
XX inhibitors of amyloid formation, deposition, accumulation and/or
XX persistence in Alzheimer's disease and other amyloidoses. The
XX CC laminin products (see AAW50888-98) may include mouse or human laminin
XX A or A1 chain, laminin B1 or B2 chain, laminin A2 chain (merosin),
XX CC laminin G1 chain, the globular repeats of the laminin A1 chain and
XX CC the beta-amyloid binding domain of the laminin A chain. A claimed
XX CC method for treating an amyloid disease comprises administering a
XX CC polypeptide having a conformational similarity to a fragment of a
XX CC laminin protein. A method for diagnosing an amyloid disease
XX involves determining levels of laminin in a sample. Production
XX of laminin or its fourth globular repeat in vivo provides a method
XX CC for in vivo inhibition of beta-amyloid amyloidosis. The products
XX and methods can be used for the diagnosis, prognosis, monitoring
XX CC and treatment of amyloidoses such as Alzheimer's disease, Down's
XX CC syndrome and hereditary cerebral haemorrhage with amyloidosis of
XX CC the Dutch type (where the specific amyloid is the beta-amyloid
XX CC protein), the amyloidosis associated with chronic inflammation,
XX CC various forms of malignancy and Familial Mediterranean Fever (AA
XX CC amyloid or inflammation-associated amyloidosis), the amyloidosis
XX CC associated with multiple myeloma and other B-cell abnormalities
XX CC (AL amyloid), the amyloidosis associated with type II diabetes
XX CC (amylin or islet amyloid), the amyloidosis associated with prion
XX CC diseases including Creutzfeldt-Jacob disease, Gertsman-Straussler
XX CC syndrome, kuru and animal scrapie (PrP amyloid), the amyloidosis
XX CC associated with long-term haemodialysis and carpal tunnel syndrome
XX CC (beta 2-microglobulin amyloid), the amyloidosis associated with
XX CC senile cardiac amyloid and Familial Amyloidotic Polynuropathy
XX CC (prealbumin or transthyretin amyloid), and the amyloidosis
XX CC associated with endocrine tumours such as medullary carcinoma of
XX CC the thyroid (variant of procalcitonin).
XX
XX Sequence 3084 AA:

```



```

; PRIOR APPLICATION NUMBER: US 60/216593
; PRIOR FILING DATE: 2000-07-07
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 15
; LENGTH: 421
; TYPE: PRN
; ORGANISM: Saccharomyces
US-09-901-252-15

Query Match
Best Local Similarity 28.3%; Score 63.5; DB 10; Length 421;
Matches 13; Conservative 11; Mismatches 11; Indels 3; Gaps 2;

QY 3 YKNRYVALKSGRLNAP--MPENGVAENNDWIFMGYT 38
Db 384 YKHFTLYLRVNGGHM-VPFDVPEALSMYENWHLGSGS 420

RESULT 3
US-09-741-669-450
; Sequence 450, Application US/09741669
; Patent No. US20020022718A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; TITLE OF INVENTION: Genes identified as required for
; FILE REFERENCE: ELITRA.009A
; CURRENT APPLICATION NUMBER: US/09/741,669
; CURRENT FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: US 60/173005
; PRIOR FILING DATE: 1999-12-23
; NUMBER OF SEQ ID NOS: 481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 450
; LENGTH: 452
; TYPE: PRN
; ORGANISM: Escherichia coli
US-09-741-669-450

Query Match
Best Local Similarity 27.7%; Score 62; DB 10; Length 452;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;

QY 12 KSGGRLNAPMPENGVA---AENNDWI 33
Db 200 KAKGEIFSGLPENGIAIMNADNDWL 225

RESULT 4
US-09-815-242-10027
; Sequence 10027, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
```

```

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10027
; LENGTH: 452
; TYPE: PRN
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(452)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-815-242-14074

Query Match
Best Local Similarity 27.7%; Score 62; DB 10; Length 452;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;

QY 12 KSGGRLNAPMPENGVA---AENNDWI 33
Db 200 KAKGEIFSGLPENGIAIMNADNDWL 225

RESULT 5
US-09-815-242-14074
; Sequence 14074, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14074
; LENGTH: 452
; TYPE: PRN
; ORGANISM: Salmonella typhi
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(452)
; OTHER INFORMATION: Xaa - Any Amino Acid
US-09-815-242-14074
```


US-09-938-275-5
; Sequence 5, Application US/09938275
; Patent No. US2002011309A1
; GENERAL INFORMATION:
; APPLICANT: Gerardo Castillo
; APPLICANT: Alan Snow
; TITLE OF INVENTION: Therapeutic and Diagnostic Applications
; TITLE OF INVENTION: of Laminin and Laminin-Derived Protein Fragments
; FILE REFERENCE: PROTEO.P03
; CURRENT APPLICATION NUMBER: US/09/938,275
; CURRENT FILING DATE: 2001-08-16
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 3075
; TYPE: PRF
; ORGANISM: Homo Sapiens
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Swissprot P25391
; DATABASE ENTRY DATE: 1992-05-01
US-09-938-275-5

Query Match 24.6%; Score 55; DB 10; Length 3075;
Best Local Similarity 30.0%; Pred. No. 32;
Matches 15; Conservative 5; Mismatches 16; Indels 14; Gaps 2;

OY 5 NYRYVALKSGR-----LNAPMPENGVAE-----NNDWIFMGTYQE 40
DB 1253 NPEPVLINGGRIRKQVITMDAPAPENGVRQEOEVAMRENFKYNVSSE 1302

RESULT 10
US-09-815-242-5118
; Sequence 5118, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5118
; LENGTH: 1317
; TYPE: PRF
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5118

Query Match 22.5%; Score 50.5; DB 10; Length 1317;
Best Local Similarity 33.3%; Pred. No. 52;

Matches 12; Conservative 4; Mismatches 9; Indels 11; Gaps 1;
OY 1 LYXKNRYVALKSGRNLNAPMPENGVAENNDWIFMG 36
DB 1122 LYXNRIRY-----LPEAGRYASODPLGLG 1146

RESULT 11
US-09-879-957-24
; Sequence 24, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; HOFFMAN, No. US20020034755A1h
; KAY, Brian K.
; FONTIKES, Dana M.
; MCCONNELL, Stephen J.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/879,957
; FILING DATE: 13-Jun-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/630,915
; FILING DATE: 03-Apr-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 1101-174
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 404 amino acids
; TYPE: amino acid
; STRANDEDNESS: <unknown>
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 24:
US-09-879-957-24
Query Match 22.3%; Score 50; DB 10; Length 404;
Best Local Similarity 52.9%; Pred. No. 16;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
OY 24 NGVAENNDWIFMGTYQE 40
DB 66 NKTAEENDLMDYHOK 82
RESULT 12
US-09-879-957-22
; Sequence 22, Application US/09879957
; Patent No. US20020034755A1
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.


```

      HOFFMAN, No. US20020034755A1H
      KAY, Brian R.
      FOLKES, Dana M.
      MCCONNELL, Stephen J.
      TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
      DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
      USING SAME
      NUMBER OF SEQUENCES: 227
      CORRESPONDENCE ADDRESS:
      ADDRESSEE: Pennie & Edmonds LLP
      STREET: 1155 Avenue of the Americas
      CITY: New York
      STATE: New York
      COUNTRY: USA
      ZIP: 10036-2711
      COMPUTER READABLE FORM:
      MEDIUM TYPE: Floppy disk
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
      SOFTWARE: Patentin Release #1.0, Version #1.30
      CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/879,957
      FILING DATE: 13-Jun-2001
      CLASSIFICATION: <Unknown>
      PRIOR APPLICATION DATA:
      APPLICATION NUMBER: US 08/630,915
      FILING DATE: 03-APR-1996
      ATTORNEY/AGENT INFORMATION:
      NAME: Mistrock, S. Leslie
      REGISTRATION NUMBER: 18,872
      REFERENCE/DOCKET NUMBER: 1101-174
      TELECOMMUNICATION INFORMATION:
      TELEPHONE: (212) 790-9090
      TELEFAX: (212) 869-8864/9741
      TELEX: 66141 PENNIE
      INFORMATION FOR SEQ ID NO: 22:
      SEQUENCE CHARACTERISTICS:
      LENGTH: 434 amino acids
      TYPE: amino acid
      STRANDEDNESS: <Unknown>
      TOPOLOGY: unknown
      MOLECULE TYPE: peptide
      SEQUENCE DESCRIPTION: SEQ ID NO: 22:
      US-09-879-957-22
      Query Match      22.3%; Score 50; DB 10; Length 434;
      Best Local Similarity 52.9%; Pred. No. 17;
      Matches      9; Conservative      4; Mismatches      4; Indels      0; Gaps      0
      OY      24 NGVAENNDWTFMGYTOE 40
      I:|||||:|:|:|
      Db      101 NKIAENNDLMDYHOK 117
      RESULT 13
      US-09-998-598-2603
      Sequence 2603, Application US/09998598
      Patent No. US20020150922A1
      GENERAL INFORMATION:
      APPLICANT: Stolck, John A.
      APPLICANT: Xu, Jiangchun
      APPLICANT: Chenaault, Ruth A.
      APPLICANT: Meagher, Madelein Joy
      TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
      TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
      FILE REFERENCE: 210121, 561
      CURRENT APPLICATION NUMBER: US/09/998,598
      CURRENT FILING DATE: 2001-11-16
      NUMBER OF SEQ ID NOS: 2606
      SOFTWARE: Corixa Invention Disclosure Database
      SEQ ID NO 2603
      LENGTH: 523
      TYPE: PRT

```

```

: ORGANISM: Homo sapiens
US-09-998-598-2603

Query Match
Best Local Similarity 55.6%; Score 48.5; DB 10; Length 523;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 5 NYRYALKSGGRINLAMP 22
   ||:|:||||| ||
Db 745 NTRFFA-QSGGRINLNP 761

RESULT 14
US-09-873-737A-6
: Sequence 6, Application US/09873737A
: Patent No. US20020076797A1
: GENERAL INFORMATION:
: APPLICANT: Duke University
: APPLICANT: Lin, Haifan
: TITLE OF INVENTION: PURIFIED AND ISOLATED PIWI FAMILY GENES AND GENE
: FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
: CURRENT FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: PCT/US99/28764
: PRIOR FILING DATE: 1999-12-03
: PRIOR APPLICATION NUMBER: 60/110,901
: PRIOR FILING DATE: 1998-12-04
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 6
: LENGTH: 861
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: (76)
: OTHER INFORMATION: Xaa=Leu or Ile
: NAME/KEY: misc_feature
: LOCATION: (303)
: OTHER INFORMATION: Xaa=Leu or Ile
: NAME/KEY: misc_feature
: LOCATION: (735)
: OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-6

Query Match
Best Local Similarity 55.6%; Score 48.5; DB 10; Length 861;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;

QY 5 NYRYALKSGGRINLAMP 22
   ||:|:||||| ||
Db 745 NTRFFA-QSGGRINLNP 761

RESULT 15
US-09-873-737A-4
: Sequence 4, Application US/09873737A
: Patent No. US20020076797A1
: GENERAL INFORMATION:
: APPLICANT: Duke University
: APPLICANT: Lin, Haifan
: TITLE OF INVENTION: PURIFIED AND ISOLATED PIWI FAMILY GENES AND GENE
: FILE REFERENCE: Attorney Docket No. US20020076797A1 180-104/2
: CURRENT FILING DATE: 2001-06-04
: PRIOR APPLICATION NUMBER: PCT/US99/28764
: PRIOR FILING DATE: 1999-12-03
: PRIOR APPLICATION NUMBER: 60/110,901
: PRIOR FILING DATE: 1998-12-04
: NUMBER OF SEQ ID NOS: 21
: SOFTWARE: PatentIn Ver. 2.1

```

```
; SEQ ID NO 4
; LENGTH: 862
; TYPE: PRT
; ORGANISM: Mus sp.
; FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (90)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (216)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (383)
;   OTHER INFORMATION: Xaa=Leu or Ile
;   NAME/KEY: misc_feature
;   LOCATION: (816)
;   OTHER INFORMATION: Xaa=Leu or Ile
US-09-873-737A-4
```

```
Query Match          21.7%; Score 48.5; DB 10; Length 862;
Best Local Similarity 55.6%; Pred. No. 61;
Matches 10; Conservative 4; Mismatches 3; Indels 1; Gaps 1;
```

```
OY 5 NYRYALKSGRLNAPMP 22
   1 1:1 :11111 1:1
Db 746 NARFFA-QSGGRLQNP 762
```

Search completed: December 10, 2002, 10:58:10
Job time : 6.2 secs

Db 1253 NPEQVLKIGRIKQVYIMDAPAPENGVOEYAMRENFVKYNSVSE 1302

RESULT 3

US-10-209-582-1019
; Sequence 1019, Application US/10209582
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM009CIN
; CURRENT APPLICATION NUMBER: US/10/209,582
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 09/758,461
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1030
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1019
; LENGTH: 62
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: misc_feature
; LOCATION: (62)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-209-582-1019

Query Match 23.0%; Score 51.5; DB 6; Length 62;

Best Local Similarity 32.4%; Pred. No. 1.8;

Matches 11; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

QY 7 RYALKSGRLNAPMPENGVAENNDWIFMGYTOE 40

Db 14 RGTTLRKGHHTASPEESGNGRME-VFLGYSFE 46

RESULT 4

US-09-134-000C-6557
; Sequence 6557, Application US/09134000C
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6557
; LENGTH: 439
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (439) (439)
; OTHER INFORMATION: Amino acid 439 is Xaa wherein Xaa = any amino acid.
US-09-134-000C-6557

Query Match 23.0%; Score 51.5; DB 5; Length 439;

Best Local Similarity 36.8%; Pred. No. 16;

Matches 14; Conservative 4; Mismatches 11; Indels 9; Gaps 1;

QY 8 YVALKSGRLNAPMPENGVAENN-----DWIFMG 36

Db 367 YVTVKSGDTLKNIAQGVSVANLRSMNGISGLIFVG 404

RESULT 5

US-10-216-209-87
; Sequence 87, Application US/10216209
; GENERAL INFORMATION:
; APPLICANT: Lam, Joseph S.
; APPLICANT: Burrows, Lori
; APPLICANT: Charter, Deborah
; APPLICANT: De Kievit, Teresa De
; TITLE OF INVENTION: Novel Proteins Involved in the Synthesis and Assembly
; FILE REFERENCE: 6580-167
; CURRENT APPLICATION NUMBER: US/10/216,209
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: US/09/352,994
; PRIOR FILING DATE: 2001-05-29
; PRIOR APPLICATION NUMBER: US 08/846,762
; PRIOR FILING DATE: 1997-04-30
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 87
; LENGTH: 378
; TYPE: PRT
; ORGANISM: Salmonella enterica
US-10-216-209-87

Query Match 22.5%; Score 50.5; DB 6; Length 378;

Best Local Similarity 33.3%; Pred. No. 19;

Matches 9; Conservative 9; Mismatches 6; Indels 3; Gaps 1;

QY 2 YKKNRYVALKSGR---LNAPMPENG 25

Db 108 FYONIKVWHIEAGLRKWNNSPPEEG 134

RESULT 6

US-10-209-582-657
; Sequence 657, Application US/10209582
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM009CIN
; CURRENT APPLICATION NUMBER: US/10/209,582
; CURRENT FILING DATE: 2002-08-01
; PRIOR APPLICATION NUMBER: 09/758,461
; PRIOR FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 1030
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 657
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (37)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-10-209-582-657

Query Match 22.3%; Score 50; DB 6; Length 82;

Best Local Similarity 36.4%; Pred. No. 4.1;

Matches 12; Conservative 5; Mismatches 10; Indels 6; Gaps 1;

QY 7 RYALKSGRLNAPMPENGVAENNDWIFMGYTO 39

Db 42 RFRVLSTNGKLNKKRPKNY-----VFKALIQ 68

```
RESULT 7
US-09-724-676-75420
; Sequence 75420, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75420
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75420

Query Match          22.3%; Score 50; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 24 NGVAENNDWIEMGYTOE 40
I :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 8
US-09-724-676-75421
; Sequence 75421, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75421
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75421

Query Match          22.3%; Score 50; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 24 NGVAENNDWIEMGYTOE 40
I :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 9
US-09-724-676A-75420
; Sequence 75420, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75420
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-75420

Query Match          22.3%; Score 50; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;
```

```
QY 24 NGVAENNDWIEMGYTOE 40
I :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 10
US-09-724-676A-75421
; Sequence 75421, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75421
; LENGTH: 329
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-75421

Query Match          22.3%; Score 50; DB 5; Length 329;
Best Local Similarity 52.9%; Pred. No. 19;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 24 NGVAENNDWIEMGYTOE 40
I :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 11
US-09-724-676-75392
; Sequence 75392, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75392
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75392

Query Match          22.3%; Score 50; DB 5; Length 365;
Best Local Similarity 52.9%; Pred. No. 21;
Matches 9; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 24 NGVAENNDWIEMGYTOE 40
I :||||| :| | | :
Db 101 NKIAENNDLMDYHOK 117

RESULT 12
US-09-724-676-75393
; Sequence 75393, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 75393
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-75393
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:50:48 ; Search time 11 seconds

(without alignments)
349,580 Million cell updates/sec

Title: US-09-142-970-5_COPY_1_40

Perfect score: 224

Sequence: 1 LYYKNRYRYALKSGGRNLNAMPENGVAENNDWIFMGYTOE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	224	100.0	1532	2 A26039	IgA-specific metal
2	190	84.8	1561	2 S61314	IgA-specific metal
3	190	84.8	1773	2 A81937	IgA-specific metal
4	186	83.0	1815	2 C81169	IgA-specific metal
5	93	41.5	1841	2 A37023	IgA-specific metal
6	93	41.5	1694	2 H64106	IgA-specific metal
7	93	41.5	1702	2 A41859	IgA-specific metal
8	85	37.9	1849	2 C41859	IgA-specific metal
9	82	36.6	1545	2 B41859	IgA-specific metal
10	63.5	28.3	532	1 CPBYX	IgA-specific metal
11	62	27.7	452	2 B90640	carboxypeptidase C
12	62	27.7	452	2 B85491	D-alanine D-alanine
13	62	27.7	452	2 A10517	UDP-N-acetyluram
14	62	27.7	452	2 F64730	UDP-N-acetyluram
15	61	27.2	297	2 S55085	hypothetical prote
16	57	25.4	189	2 B69020	hypothetical prote
17	57	25.4	1367	2 T33819	hypothetical prote
18	55	24.8	3084	1 MMMSA	laminin alpha-1 ch
19	55	24.8	766	2 T20003	hypothetical prote
20	55.5	24.8	993	2 A38437	probable homeotic
21	55	24.6	574	2 A91182	hypothetical prote
22	55	24.6	574	2 E86028	hypothetical prote
23	55	24.6	658	2 E95111	endo-beta-N-acetyl
24	55	24.6	721	2 C97980	endo-beta-N-acetyl
25	54	24.6	3075	2 S14458	laminin alpha-1 ch
26	54	24.1	350	2 S47292	phenol 2-monooxyge
27	53.5	23.9	660	2 S70904	transferrin-bindin
28	53.5	23.9	671	2 A38109	autolysin - Entero
29	53.5	23.9	850	2 S20462	RNA12 protein - ye

30	53	23.7	1076	2 A69409	carbamoyl-phosphat
31	52.5	23.4	949	2 H97322	DNA/RNA helicase,
32	52	23.2	262	2 A25140	paraportal crystal
33	52	23.2	431	2 S50633	hypothetical prote
34	52	23.2	530	2 T35201	probable transmemb
35	52	23.2	577	2 AD1440	hypothetical prote
36	52	23.2	821	2 C64461	hypothetical prote
37	51.5	23.0	349	2 T41892	BRO-d orf2 - Bomby
38	51.5	23.0	447	2 S37048	cysteine proteinas
39	51.5	23.0	507	2 S50949	hypothetical prote
40	51.5	23.0	4967	2 S72269	ryanodine receptor
41	51.5	23.0	4969	2 A37113	ryanodine receptor
42	51	22.8	641	2 A55549	glucan 1,6-alpha-1
43	51	22.8	8243	2 T31307	type I fatty acid
44	50.5	22.5	375	2 A83788	spore germination
45	50.5	22.5	472	2 B86730	dipeptidase [impor

ALIGNMENTS

RESULT 1

A26039

IgA-specific metalloendopeptidase (EC 3.4.24.13) precursor - Neisseria gonorrhoeae (s
N:Alternate names: IgA protease; immunoglobulin A1 proteinase

C:Species: Neisseria gonorrhoeae

A:Variety: strain MS11

C:Date: 05-Oct-1988 #sequence_revision 05-Oct-1988 #text_change 08-Dec-2000

C:Accession: A26039; S09386

R:Polmer, J.; Halter, R.; Beyreuther, K.; Meyer, T.F.

Nature 325, 458-462, 1987

A:Title: Gene structure and extracellular secretion of Neisseria gonorrhoeae IgA prot

A:Reference number: A26039; PMID:87115823; PMID:3027577

A:Accession: A26039

A:Molecule type: DNA

A:Residues: 1-1532 <POH>

A:Cross-references: GB:X04835; NID:944868; PIDN:CAA28538.1; PID:944869

A:Note: the authors translated the codon AAG for residue 668 as Asn

R:Halter, R.; Polmer, J.; Meyer, T.F.

EMBO J 8, 2737-2744, 1989

A:Title: Mosaic-like organization of IgA protease genes in Neisseria gonorrhoeae gene

A:Reference number: S09386; PMID:90060036; PMID:2511009

A:Accession: S09386

A>Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 281-325, 'N', 327-337, 'N', 339-427, 'W', 429-531, 'N', 533-615, 'V', 617-631, 'N', 6

A:Experimental source: strain MS11

C:Genetics:

A:Gene: IgA

C:Keywords: hydrolyase; metalloprotease; transmembrane protein

F:127/Domain: signal sequence #status predicted <Sig>

F:28-1532/Product: immunoglobulin A1 proteinase #status predicted <Mat>

F:986-987/Cleavage site: Pro-Ser (autolytic) #status predicted

F:1018-1019/Cleavage site: Pro-Ser (autolytic) #status predicted

F:1121-1122/Cleavage site: Pro-Ala (autolytic) #status predicted

Query Match 100.0%; Score 224; DB 2; Length 1532;
Best Local Similarity 100.0%; Pred. No. 7,4e-22;

Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYYKNRYRYALKSGGRNLNAMPENGVAENNDWIFMGYTOE 40

|||||

DB 584 LYYKNRYRYALKSGGRNLNAMPENGVAENNDWIFMGYTOE 623

|||||

RESULT 2

S61314

IgA-specific metalloendopeptidase (EC 3.4.24.13) homolog - Neisseria meningitidis

C:Species: Neisseria meningitidis

A:Variety: Hfl3

C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 08-Dec-2000

C:Accession: S61314

R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IgA1 protease in *Neisseria*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1561 <LOW>
A:Cross-references: EMBL:X82474; NID:g732873; PIDN:CAA57857.1; PID:g732874
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 84.8%; Score 190; DB 2; Length 1561;
Best Local Similarity 85.0%; Pred. No. 3.1e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWTFMGYTOE 40
|||||
Db 584 LYYKNRYALKSGSVNAPMPENGOTENNNDWLMGSTOE 623

RESULT 3

IGA-specific metalloendopeptidase (EC 3.4.24.13) NMA0905 [imported] - *Neisseria meningitidis*
N:Alternate names: Iga1 proteinase; Iga1 proteinase (EC 3.4.21.7) [misnomer]; immunoglobulin
C:Species: *Neisseria meningitidis*
A:Variety: strain Z2491 serogroup A; strain HF117; strain HF159; strain SM1027
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: A81937; S61317; S61318; S61321
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtzoy, S.; Jørgensen, K.; Leather, S.; Mølle, S.; Mungall, K.; Quail, M.A.; Rajandream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.
A:Reference number: A81775; MUID:2022556; PMID:10761919
A:Accession: A81937
A:Molecule type: DNA
A:Residues: 1-1773 <PAR>
A:Cross-references: GB:AL162754; GB:AL157959; NID:g7379424; PIDN:CA884182.1; PID:g737961
A:Experimental source: serogroup A, strain Z2491
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IgA1 protease in *Neisseria*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61317
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82470; NID:g732854; PIDN:CAA57853.1; PID:g732855
A:Experimental source: strain HF117
A:Accession: S61318
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82471; NID:g732858; PIDN:CAA57854.1; PID:g732859
A:Experimental source: strain HF159
A:Accession: S61321
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82472; NID:g732852; PIDN:CAA57855.1; PID:g732853
A:Experimental source: strain SM1027
C:Genetics:
A:Gene: iga; NMA0905
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 84.8%; Score 190; DB 2; Length 1773;
Best Local Similarity 85.0%; Pred. No. 3.6e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWTFMGYTOE 40
|||||
Db 573 LYYKNRYALKSGSVNAPMPENGOTENNNDWLMGSTOE 612

RESULT 4

IGA-specific metalloendopeptidase (EC 3.4.24.13) NMB0700 [imported] - *Neisseria meningitidis*
C:Species: *Neisseria meningitidis*
A:Variety: strain MC58 serogroup B; strain 81139
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: C81169; S61326
R:Rettelin, H.; Sanders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Half, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
; H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignan, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappelli, R.;
A:Title: Complete genome sequence of *Neisseria meningitidis* serogroup B strain MC58.
A:Reference number: A81000; MUID:20175755; PMID:10710307
A:Accession: C81169
A:Molecule type: DNA
A:Residues: 1-1815 <RET>
A:Cross-references: GB:AE002424; GB:AE002098; NID:g7225923; PIDN:AAF41117.1; PID:g722
A:Experimental source: serogroup B, strain MC58
R:Lomholt, H.; Poulsen, K.; Mogens, K.
Mol. Microbiol. 15, 495-506, 1995
A:Title: Comparative characterization of the iga gene encoding IgA1 protease in *Neiss*
A:Reference number: S61314; MUID:95302961; PMID:7783620
A:Accession: S61326
A:Status: preliminary; nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 53-548 <LOW>
A:Cross-references: EMBL:X82477; NID:g732856; PIDN:CAA57860.1; PID:g732857
A:Genetics:
A:Gene: NMB0700
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 83.0%; Score 186; DB 2; Length 1815;
Best Local Similarity 82.5%; Pred. No. 1.3e-16;
Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWTFMGYTOE 40
|||||
Db 573 LYYKNRYALKSGSVNAPMPENGOTENNNDWLMGSTOE 612

RESULT 5

IGA-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - *Haemophilus influenzae*
N:Alternate names: immunoglobulin A1 proteinase type 1
C:Species: *Haemophilus influenzae*
C:Date: 31-Jan-1992 #sequence_revision 12-Jun-1992 #text_change 08-Dec-2000
C:Accession: A37023
R:Poulsen, K.; Brandt, J.; Hjorth, J.P.; Thøgersen, H.C.; Killian, M.
Infect. Immun. 57, 3097-3105, 1989
A:Title: Cloning and sequencing of the immunoglobulin A1 protease gene (iga) of *Haemo*
A:Reference number: A37023; MUID:89379374; PMID:2506130
A:Accession: A37023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1541 <POU>
A:Cross-references: GB:X64357; NID:g433560; PIDN:CAA45708.1; PID:g433561
A:Experimental source: serotype b
C:Superfamily: Iga-specific metalloendopeptidase
C:Keywords: hydrolase; metalloproteinase

Query Match 41.5%; Score 93; DB 2; Length 1541;
Best Local Similarity 37.5%; Pred. No. 0.00044;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

QY 1 LYYKNRYALKSGRLNAPMPENGVAENNDWTFMGYTOE 40
|||||
Db 595 LNEYTYTALRKAGSTRSELPKNSGSENNWLMGSTSD 634

RESULT 6

H64106

IGa-specific metalloendopeptidase (EC 3.4.24.13) type 1 - Haemophilus influenzae (strain N:Alternante names: immunoglobulin A1 proteolase type 1

C:Species: Haemophilus influenzae

A:Variety: strain Rd KW20

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Dec-2000

A:Accession: H64106; A41500

R:Flensburg, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.

; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geophagen, N.S.M.

Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: H64106

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1694 <TIGR>

A:Cross-references: GB:U32779; GB:I42023; NID:91574009; PIDN:AA22651.1; PID:91574019; T

A:Experimental source: strain Rd KW20

R:Grundy, F.J.; Plaut, A.G.; Wright, A.

Infect. Immun. 58, 320-331, 1990

A:Title: Localization of the cleavage site specificity determinant of Haemophilus influe

A:Reference number: A41500; MUID:90129281; PMID:2105270

A:Accession: A41500

A:Molecule type: DNA

A:Residues: 1-377 <GRD>

A:Cross-references: GB:X59800

A:Experimental source: strain Rd KW20

A:Note: the authors translated the codon TCG for residue 319 as Thr

C:Function:

A:Description: this protease is classified as type 1 because it cleaves at a proline-s

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

Query Match

Best Local Similarity 41.5%; Score 93; DB 2; Length 1694;

Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGRLNMPENGVAENNDFMGCTOE 40

DB 601 LNFENRYVALKSGRLNMPENGVAENNDFMGCTSD 640

RESULT 7

A41859

IGa-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus influe

C:Species: Haemophilus influenzae

A:Variety: strain HK715

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: A41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae ty

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: A41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1702 <POU>

A:Cross-references: GB:M87489; NID:9148906; PIDN:AAA24966.1; PID:9148907

A:Experimental source: strain HK715

A:Note: sequence extracted from NCBI backbone (NCBIP:97282)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

Query Match

Best Local Similarity 41.5%; Score 93; DB 2; Length 1702;

Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYYKNRYVALKSGRLNMPENGVAENNDFMGCTOE 40

DB 601 LNFENRYVALKSGRLNMPENGVAENNDFMGCTSD 640

RESULT 8

C41859

IGa-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK613

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: C41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: C41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1849 <POU>

A:Experimental source: strain HK613

A:Note: sequence extracted from NCBI backbone (NCBIP:97285)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

Query Match

Best Local Similarity 37.9%; Score 85; DB 2; Length 1849;

Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

OY 1 LYYKNRYVALKSGRLNMPENGVAENNDFMGCTOE 40

DB 604 LNFENRYVALKSGRLNMPENGVAENNDFMGCTSD 645

RESULT 9

B41859

IGa-specific metalloendopeptidase (EC 3.4.24.13) type 1 precursor - Haemophilus infl

C:Species: Haemophilus influenzae

A:Variety: strain HK393

C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 08-Dec-2000

C:Accession: B41859

R:Poulsen, K.; Reinholdt, J.; Kilian, M.

J. Bacteriol. 174, 2913-2921, 1992

A:Title: A comparative genetic study of serologically distinct Haemophilus influenzae

A:Reference number: A41859; MUID:92234949; PMID:1373717

A:Accession: B41859

A:Status: preliminary; not compared with conceptual translation

A:Molecule type: nucleic acid

A:Residues: 1-1545 <POU>

A:Cross-references: GB:M87490; NID:9148908; PIDN:AAA24967.1; PID:9148909

A:Experimental source: strain HK393

A:Note: sequence extracted from NCBI backbone (NCBIP:97283)

C:Superfamily: IGA-specific metalloendopeptidase

C:Keywords: hydrolase; metalloprotease

Query Match

Best Local Similarity 36.6%; Score 82; DB 2; Length 1545;

Matches 15; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

OY 1 LYYKNRYVALKSGRLNMPENGVAENNDFMGCTOE 36

DB 597 LNFENRYVALKSGRLNMPENGVAENNDFMGCTSD 634

RESULT 10

CPRY

carboxypeptidase C (EC 3.4.16.5) precursor [validated] - yeast (Saccharomyces cerevis

N:Alternante names: carboxypeptidase Y; protein YIM297w

C:Species: Saccharomyces cerevisiae

C:Date: 15-Oct-1982 #sequence_revision 31-Mar-1993 #text_change 15-Sep-2000

C:Accession: A26597; S47458; A90763; A94609; A00909

R:Valls, L.A.; Hunter, C.P.; Rothman, J.H.; Stevens, T.H.

Cell 48, 887-897, 1987

A:Title: Protein sorting in yeast: the localization determinant of yeast vacuolar car

A:Reference number: A26597; MUID:87131100; PMID:3028649

A:Accession: A26597

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:17:38 ; Search time 5.6 seconds

(without alignments)
296.259 Million cell updates/sec

Title: US-09-142-970-5_COPY_1_40

Perfect score: 224

Sequence: 1 LYKKNRYALKSGRLNAPMPENGVAMNDWIFMGTYQE 40

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	224	100.0	1532	1	IGA_NEIGO
2	93	41.5	1541	1	IGA1_HAEIN
3	93	41.5	1694	1	IGA2_HAEIN
4	93	41.5	1702	1	IGA2_HAEIN
5	85	37.9	1849	1	IGA3_HAEIN
6	82	36.6	1545	1	IGA3_HAEIN
7	63.5	26.3	532	1	CBPY_YEAST
8	62	27.7	452	1	MURF_YEAST
9	61	27.2	297	1	MYR9_YEAST
10	57	25.4	3084	1	LMAL_MOUSE
11	55.5	24.8	993	1	TSH_DROME
12	55	24.6	658	1	LYTB_STRPN
13	55	24.6	3075	1	LMAL_HUMAN
14	53.5	22.9	671	1	ALYS_ENTEA
15	53.5	22.9	850	1	RNI2_YEAST
16	53	23.7	1076	1	CARB_ARCFU
17	52	23.2	443	1	YEM0_YEAST
18	51.5	23.0	507	1	YME3_YEAST
19	51.5	23.0	4967	1	RVR2_HUMAN
20	51.5	23.0	4969	1	RVR2_RABIT
21	51	22.8	641	1	IMD_ARTGO
22	50.5	22.5	378	1	RFCB_SALBO
23	50.5	22.5	1361	1	RPOD_SPIOL
24	50	22.3	563	1	YHWI_ECOLI
25	50	22.3	588	1	BINI_MOUSE
26	50	22.3	588	1	BINI_MOUSE
27	50	22.3	593	1	BINI_MOUSE
28	50	22.3	1193	1	DP3A_XYLLA
29	49.5	22.1	520	1	Y120_MYCGE
30	49.5	22.1	704	1	GIS2_YEAST
31	49	21.9	100	1	YF78_MYCPN
32	49	21.9	359	1	CD72_HUMAN
33	49	21.9	801	1	DHGA_ACICA

34	48.5	21.7	114	1	YEOG_ECOLI
35	48.5	21.7	361	1	SEP2_HUMAN
36	48.5	21.7	361	1	SEP2_MOUSE
37	48.5	21.7	707	1	UTP1_MOUSE
38	48.5	21.7	708	1	UTP1_MOUSE
39	48.5	21.7	1032	1	VG07_BPT4
40	48.5	21.7	1225	1	Y309_MYCGE
41	48	21.4	106	1	Y086_CAEEL
42	48	21.4	296	1	ALYS_BDDP1
43	48	21.4	394	1	VORA_PRRHO
44	48	21.4	668	1	TKT1_PASMO
45	48	21.4	668	1	TKT2_PASMO

ALIGNMENTS

RESULT 1	IGA_NEIGO	STANDARD:	PRT: 1532 AA.
ID	IGA_NEIGO		
AC	P09790:		
DT	01-MAR-1989 (rel. 10, Created)		
DT	01-MAR-1989 (rel. 10, Last sequence update)		
DT	15-JUN-2002 (rel. 41, Last annotation update)		
DE	IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA protease).		
GN	IGA.		
OS	Neisseria gonorrhoeae.		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_Taxid=485;		
RN	[1]		
RP	SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.		
RC	STRAIN=MS11;		
RA	MEDLINE=87115823; PubMed=3027577;		
RT	Pohlner J., Halter R., Beyreuther K., Meyer T.F.;		
RT	"Gene structure and extracellular secretion of Neisseria gonorrhoeae Hemophilus influenzae by peptide prolyl boronic acids.";		
RL	J. Biol. Chem. 265:3738-3743(1990).		
CC	- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.		
CC	- CATALYTIC ACTIVITY: cleavage of immunoglobulin A molecules at certain Pro-I-Xaa bonds in the hinge region. No small molecule substrates are known.		
CC	- SUBCELLULAR LOCATION: Secreted.		
CC	- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE		
CC	DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.		
CC	- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL: X04835; CAA28538.1; -		P33917 escherichia
DR	PIR: A26039; A26039.		Q15019 homo sapien
DR	MEROPS: S06.001; -		P42208 mus musculu
DR	InterPro: IPR000710; IGA_S6.		Q9w4i9 mus musculu
DR	InterPro: IPR004899; Pertactin_sup.		Q9r237 r c-jun-aml
DR	Pfam: PF02395; IGA1; 1.		P19061 bacterioph
DR	Pfam: PF03212; Pertactin; 1.		P47551 mycoplasma
DR	PRINTS: PR00921; IGASERPTASE.		009238 caenorhabd
DR	Hydrolase; Serine protease; zymogen; Autocatalytic cleavage;		003979 bacterioph
			O58413 pyrococcus
			P57927 pasteurella
			P57958 pasteurella

```

KW Transmembrane; Signal.
FT SIGNAL 1 27
FT CHAIN 28 986 IGA-SPECIFIC SERINE ENDOPEPTIDASE.
FT PROPEP 987 1532 HELPER PEPTIDE.
FT ACT_SITE 278 278 POTENTIAL.
FT SITE 986 987 CLEAVAGE (AUTO-).
FT SITE 1018 1019 CLEAVAGE (AUTO-).
FT SITE 1121 1122 CLEAVAGE (AUTO-).
SQ SEQUENCE 1532 AA; 168976 MW; 68FF4112BD22F40D CRC64;

Query Match
Best Local Similarity 100.0%; Score 224; DB 1; Length 1532;
Pred. No. 1; Ie-22;
Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LYKNRYTALKSGRLNAPMPENGVAENNDWIFMGCTQE 40
Db 584 LYKNRYTALKSGRLNAPMPENGVAENNDWIFMGCTQE 623

RESULT 2
IGA0_HAEIN STANDARD; PRT; 1541 AA.
AC P42782;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK368 / Serotype B;
RX MEDLINE=89379374; PubMed=2506130;
RA Poulsen K., Brandt J., Hjorth J.P., Thøgersen H.C., Killian M.;
RT "Cloning and sequencing of the immunoglobulin A1 protease gene (iga)
RT of Haemophilus influenzae serotype b."
RL Infect. Immun. 57:3097-3105(1989).
RN [2]
RP MUTAGENESIS OF SER-288.
RC STRAIN=HK368 / Serotype B;
RX MEDLINE=92234949; PubMed=1373117;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases."
RL J. Bacteriol. 174:2913-2921(1992).
CC -I- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: X64357; CAA45708.1; -
CC DR EMBL: M87492; AAA24969.1; -
CC DR MEROPS: S06.001; -
CC InterPro: IPR000710; IGA_S6.

```

```

DR InterPro: IPR004899; Pertactin_sup.
DR Pfam: PF02395; IGA1_1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PRO0921; IGASERPTASE.
KW Hydrolase; Serine protease; Transmembrane; Zymogen; Signal.
FT SIGNAL 1 25
FT CHAIN 26 1008 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1009 1541 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
SQ SEQUENCE 1541 AA; 169370 MW; CE7257CB3196C600 CRC64;

Query Match
Best Local Similarity 41.5%; Score 93; DB 1; Length 1541;
Pred. No. 0.00013;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

OY 1 LYKNRYTALKSGRLNAPMPENGVAENNDWIFMGCTQE 40
Db 595 INLENTYALRKGASTRSELPRNKGSENMWLYMGRTSD 634

RESULT 3
IGA0_HAEIN STANDARD; PRT; 1694 AA.
AC P44969;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA OR IGA1 OR HI0990.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Serotype D;
RX Wright A., Fishman Y., Tai F., Plaut A.G.;
RL Submitted (MAR-1991) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Fierman J., Gocayne J.D., Lasker J.M., et al.;
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.B., Shirley R., Liu L.T., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Heddlom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -I- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC PRODUCING INTRACT FC AND FAB FRAGMENTS.
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entries requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X59800: -; NOT_ANNOTATED_CDS.
DR EMBL: U32779: AAC22651.1: -.
DR MEROPS: S06.001: -.
DR TIGR: H10990: -.
DR InterPro: IPR000710: IGA_S6.
DR InterPro: IPR004899: Pertactin_sup.
DR Pfam: PF02395: IGA1; 1.
DR Pfam: PF03212: Pertactin; 2.
DR PRINTS: PR00921: IGASERPRASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; signal;
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1694 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT CONFLICT 253 254 EN -> GV (IN REF. 1).
FT CONFLICT 272 272 G -> A (IN REF. 1).
FT CONFLICT 464 464 G -> E (IN REF. 1).
FT CONFLICT 866 866 S -> T (IN REF. 1).
FT CONFLICT 1036 1036 A -> D (IN REF. 1).
FT CONFLICT 1074 1074 A -> G (IN REF. 1).
FT CONFLICT 1421 1421 A -> G (IN REF. 1).
FT CONFLICT 1545 1545 H -> T (IN REF. 1).
SQ SEQUENCE 1694 AA; 185539 MW; C52427013F93176C CRC64;

QY 1 LYYKNRYVALKSGRLNMPENGVAENNDWIFMGTYOE 40
Best Local Similarity 41.5%; Score 93; DB 1; Length 1694;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

DB 601 LNLENTTYALRKAGSTRSELPRKSGSENNWLYMGKXTSD 640

RESULT 4
IGA2_HAEIN
ID IGA2_HAEIN STANDARD; PRT; 1702 AA.
AC P45384;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK715 / Serotype B;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-1-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M87489: AAA24966.1: -.
DR MEROPS: S06.001: -.
DR InterPro: IPR000710: IGA_S6.
DR InterPro: IPR004899: Pertactin_sup.
DR Pfam: PF02395: IGA1; 1.
DR Pfam: PF03212: Pertactin; 2.
DR PRINTS: PR00921: IGASERPRASE.
KW Hydrolyase; Serine protease; Transmembrane; zymogen; Repeat; signal.
KW Complete proteome.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 1014 IMMUNOGLOBULIN A1 PROTEASE.
FT PROPEP 1015 1702 HELPER PEPTIDE (POTENTIAL).
FT ACT_SITE 288 288 PROBABLE.
FT DOMAIN 1109 1124 2 X 8 AA TANDEM REPEATS OF A-K-V-E-K-E-E-
FT REPEAT 1109 1116 K.
FT REPEAT 1117 1124 2.
SQ SEQUENCE 1702 AA; 186539 MW; 860F70D267807A6 CRC64;

QY 1 LYYKNRYVALKSGRLNMPENGVAENNDWIFMGTYOE 40
Best Local Similarity 37.5%; Score 93; DB 1; Length 1702;
Matches 15; Conservative 9; Mismatches 16; Indels 0; Gaps 0;

DB 601 LNLENTTYALRKAGSTRSELPRKSGSENNWLYMGKXTSD 640

RESULT 5
IGA4_HAEIN
ID IGA4_HAEIN STANDARD; PRT; 1849 AA.
AC P45386;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).
GN IGA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_Taxid=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NH1 HK61;
RC MEDLINE=92234949; PubMed=1373717;
RA Poulsen K., Reinholdt J., Killian M.;
RT "A comparative genetic study of serologically distinct Haemophilus
RT influenzae type 1 immunoglobulin A1 proteases.";
RT J. Bacteriol. 174:2913-2921(1992).
CC -1- FUNCTION: VIRULENCE FACTOR; CLEAVES HOST IMMUNOGLOBULIN A
CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-1-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY
CC SIMILARITY).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch.

CC -----

DR EMBL: M87491; AAA24968.1; -

DR MEROPS: S06.001; -

DR InterPro: IPR000710; IGA_S6.

DR InterPro: IPR004899; Pertactin_sup.

DR Pfam: PF03295; IGA1; 1.

DR Pfam: PF03212; Pertactin; 2.

DR PRINTS: PR00921; IGASERPTASE.

DR HydroLase: Serine protease; Transmembrane; zymogen; Signal.

FT CHAIN 1 25

FT SIGNAL 1 25

FT PROPEP 1022 1849 IMMUNOGLOBULIN A1 PROTEASE.

FT ACT_SITE 299 299 HELPER PEPTIDE (POTENTIAL).

FT ACT_SITE 299 299 PROBABLE.

SQ SEQUENCE 1849 AA; 202957 MW; 79A7D018C7150AEA CRC64;

Query Match 37.9%; Score 85; DB 1; Length 1849;

Best Local Similarity 35.7%; Pred. No. 0.002;

Matches 15; Conservative 8; Mismatches 17; Indels 2; Gaps 1;

QY 1 LYRK--NYRYALKSGGRINAMPENGVAENNDWIFMGCTOE 40

DB 604 LYFQNDNRSYTLKKKGASTRSLPQNSGSESNEMLYMGRTSD 645

RESULT 6

IGA3_HAEIN STANDARD; PRT; 1545 AA.

AC P45385;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Immunoglobulin A1 protease precursor (EC 3.4.21.72) (IGA1 protease).

GN IGA.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

NCBI_TaxID=727;

ON [1]

RP SEQUENCE FROM N.A.

RC STRAIN=HK393 / NCTC 8467 / Serotype B;

RX MEDLINE=92234949; PubMed=1373717;

RA Poulsen K., Reinholdt J., Kilian M.;

RT "A comparative genetic study of serologically distinct Haemophilus influenzae type 1 immunoglobulin A1 proteases.";

RT J. Bacteriol. 174:2913-2921(1992).

CC -1- FUNCTION: VIRULENCE FACTOR: CLEAVES HOST IMMUNOGLOBULIN A PRODUCING INTACT FC AND FAB FRAGMENTS.

CC -1- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at certain Pro-I-Xaa bonds in the hinge region. No small molecule substrates are known.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- DOMAIN: THE SIGNAL PEPTIDE GUIDES THE PRECURSOR TO THE PERIPLASMIC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: M87490; AAA24967.1; -

DR MEROPS: S06.001; -

DR InterPro: IPR000710; IGA_S6.

DR InterPro: IPR004899; Pertactin_sup.

DR Pfam: PF03295; IGA1; 1.

DR Pfam: PF03212; Pertactin; 2.

DR PRINTS: PR00921; IGASERPTASE.

KW HydroLase; Serine protease; Transmembrane; zymogen; Signal.

FT SIGNAL 1 25

FT CHAIN 26 1012 IMMUNOGLOBULIN A1 PROTEASE.

FT PROPEP 1013 1545 HELPER PEPTIDE (POTENTIAL).

FT ACT_SITE 292 292 PROBABLE.

SQ SEQUENCE 1545 AA; 170627 MW; 3EDD753988F6D478 CRC64;

Query Match 36.6%; Score 82; DB 1; Length 1545;

Best Local Similarity 39.5%; Pred. No. 0.0042;

Matches 15; Conservative 7; Mismatches 14; Indels 2; Gaps 1;

QY 1 LYK--KNRYRYALKSGGRINAMPENGVAENNDWIFMG 36

DB 597 LYFNEENRYTALKKQASIRSEFPQNSGSESNMGLYMG 634

RESULT 7

CBPY_YEAST STANDARD; PRT; 532 AA.

AC P00729;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Carboxypeptidase Y precursor (EC 3.4.16.5) (Carboxypeptidase YSCY).

GN PC1 OR YMR297W.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

NCBI_TaxID=4932;

ON [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=87131100; PubMed=3028649;

RA Valls L.A., Hunter C.P., Rothman J.H., Stevens T.H.;

RT "Protein sorting in yeast: the localization determinant of yeast vacuolar carboxypeptidase Y resides in the propeptide.";

RT Cell 48:887-897(1987).

RL [2]

RP SEQUENCE FROM N.A.

RC STRAIN=S288c / AB972;

RX Badcock K., Churcher C., Barrell B.G., Rajandream M.A.;

RL Submitted (MIG-1994) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE OF 112-532.

RA Svendsen I., Martin B.M., Viswanatha T., Johansen J.T.;

RT "Amino acid sequence of carboxypeptidase Y. II. Peptides from RT enzymatic cleavages.";

RL Carlsberg Res. Commun. 47:15-27(1982).

RN [4]

RP REVISIONS, AND ACTIVE SITE SER-257.

RA Bredam K., Svendsen I.;

RT "Identification of methionyl and cysteinyl residues in the substrate RT binding site of carboxypeptidase Y.";

RL Carlsberg Res. Commun. 49:639-645(1984).

RN [5]

RP ACTIVE SITE HIS-508.

RX MEDLINE=90315013; PubMed=2639680;

RA Beech L.M., Bredam K.;

RT "Inactivation of carboxypeptidase Y by mutational removal of the putative essential histidyl residue.";

RL Carlsberg Res. Commun. 54:165-171(1989).

RN [6]

RP MUTAGENESIS.

RX MEDLINE=94114535; PubMed=7904479;

RA Mortensen U.H., Remington S.J., Bredam K.;

RT "Site-directed mutagenesis on (serine) carboxypeptidase Y. A hydrogen bond network stabilizes the transition state by interaction with the C-terminal carboxylate group of the substrate.";

RL Biochemistry 33:508-517(1994).

RN [7]

RP X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).

RX MEDLINE=95244421; PubMed=7727362;

RA Endrizzi J.A., Bredam K., Remington S.J.;

RT "2.8-A structure of yeast serine carboxypeptidase.";
RL Biochemistry 33:1106-1120(1994).
CC -1- FUNCTION: INVOLVED IN DEGRADATION OF SMALL PEPTIDES. DIGESTS
CC PREFERENTIALLY PEPTIDES CONTAINING AN ALIPHATIC OR HYDROPHOBIC
CC RESIDUE IN P1' POSITION. AS WELL AS METHIONINE, LEUCINE OR
CC PHENYLALANINE IN P1 POSITION OF ESTER SUBSTRATE.
CC -1- CATALYTIC ACTIVITY: Release of a C-terminal amino acid with a
CC broad specificity.
CC -1- ENZYME REGULATION: INHIBITED BY 2PCK.
CC -1- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
CC -1- PTM: ENTERS THE ENDOPLASMIC RETICULUM AS AN INACTIVE ZMOGEN AND
CC IS MODIFIED BY FOUR N-LINKED CORE OLIGOSACCHARIDES, GIVING RISE TO
CC A PRECURSOR KNOWN AS P1 (67 kDa). AS P1 TRANSITS THROUGH THE
CC GOLGI, EXTENSION OF ITS CORE OLIGOSACCHARIDES LEADS TO THE
CC GOLDI-MODIFIED P2 PRECURSOR (69 kDa). P2 IS SORTED AWAY FROM
CC SECRETORY PROTEINS AT OR BEYOND A LATE GOLGI COMPARTMENT AND IS
CC SUBSEQUENTLY DELIVERED TO THE VACUOLE VIA A PREVACUOLAR
CC ENDOSOME-LIKE COMPARTMENT. UPON ARRIVAL IN THE VACUOLE, THE
CC N-TERMINAL PROSEGMENT OF P2 IS CLEAVED TO YIELD THE ENZYMATICALLY
CC ACTIVE MATURE VACUOLAR FORM OF CPY (61 kDa).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 10.
CC -1- DATABASE: NAME=worthington enzyme manual;
CC WWW="http://www.worthington-biochem.com/manual/C/COY.html".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15482; AAA34902.1; -;
DR EMBL: X80836; CAA56806.1; -;
DR PIR: A26597; CPHYX.
DR PDB: 1YSC; 22-JUN-94.
DR PDB: 1CPY; 15-SEP-95.
DR MEROPS: S10.001; -;
DR SGD: S0004912; PRCL.
DR InterPro: IPR000379; Ser-estrs-site.
DR InterPro: IPR001563; Serine-carbpept.
DR Pfam: PF00450; serine-carbpept; 1.
DR PRINTS: PR00724; CRBOXYPTASEC.
DR PRODOM: PD001189; Serine-carbpept; 1.
DR PROSITE: PS00131; CARBOXYPEPT_SER_SER; 1.
DR PROSITE: PS00560; CARBOXYPEPT_SER_HIS; 1.
KW Hydrolase; Carboxypeptidase; Glycoprotein; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 20
FT PROPEP 21 111
FT CHAIN 112 532
FT ACT_SITE 257 257
FT ACT_SITE 449 449
FT ACT_SITE 508 508
FT BINDING 452 452
FT BINDING 509 509
FT BINDING 552 552
FT DISULFID 167 409
FT DISULFID 304 318
FT DISULFID 328 351
FT DISULFID 335 344
FT DISULFID 373 379
FT DISULFID 124 124
FT CARBOHYD 198 198
FT CARBOHYD 279 279
FT CARBOHYD 479 479
FT SITE 24 27
FT MUTAGEN 508 508
FT CONFLICT 260 261
FT CONFLICT 389 389
FT CONFLICT 529 529
FT SEQUENCE 532 AA; 59802 MW; 7227F3489CDD9552 CRC64;
Query Match 28.3%; Score 63.5; DB 1; Length 532;

Best Local Similarity 34.2%; Pred. No. 0.44;
Matches 13; Conservative 11; Mismatches 11; Indels 3; Gaps 2;
QY 3 YKNRYRYALKSGGRINAP--MPENGAENNDFWFMCT 38
DB 495 YKHFTYLRVFNQGHM-VPEDVPENALSMVNEWTHGFS 531
RESULT 8
MURF_ECOLI
ID MURF_ECOLI STANDARD; PRT; 452 AA.
AC P11860; P77636; 007100;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE UDP-N-acetylglucosaminylalanyl-D-glutimyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNac-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MURF OR MRA OR B0086.
OS Escherichia coli.
OC Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxId=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-89345095; PubMed-2668880;
RA Parquet C., Flouret B., Mengin-Lecreulx D., van Heijenoort J.;
RT "Nucleotide sequence of the murF gene encoding the UDP-MurNac-
RT pentapeptide synthetase of Escherichia coli.";
RL Nucleic Acids Res. 17:5379-5379(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-92334977; PubMed-1630901;
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-MB2884;
RX MEDLINE-97128642; PubMed-8973200;
RA Anderson M.S., Eveland S.S., Onishi H.R., Pompliano D.L.;
RT "Kinetic mechanism of the Escherichia coli UDPMurNac-tripeptide
RT D-alanyl-D-alanine-adding enzyme: use of a glutathione S-transferase
RT fusion.";
RL Biochemistry 35:16264-16269(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP SEQUENCE FROM N.A., AND MUTANT MURF2.
RC STRAIN-CGSC 5990;
RX MEDLINE-97309380; PubMed-9166795;
RA Eveland S.S., Pompliano D.L., Anderson M.S.;
RT "Conditionally lethal Escherichia coli murEIN mutants contain point
RT defects that map to regions conserved among murEIN and folYI poly-
RT gamma-glutamate ligases: identification of a ligase superfamily.";
RL Biochemistry 36:6223-6229(1997).
RN [6]
RP CHARACTERIZATION, AND SEQUENCE OF 1-15.
RX MEDLINE-90248455; PubMed-2186811;
RA Duncan K., van Heijenoort J., Walsh C.T.;
RT "Purification and characterization of the D-alanyl-D-alanine-adding

```

RT enzyme from Escherichia coli."
RL Biochemistry 29:2379-2386(1990).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).
RX MEDLINE=20545602; PubMed=11090285;
RA Yan Y., Munshi S., Leitling B., Anderson M.S., Chirvas J., Chen Z.;
RT "Crystal structure of Escherichia coli UDPMurNAc-tripeptide
RL D-alanyl-D-alanine-adding enzyme (MurF) at 2.3-A resolution.";
RL J. Mol. Biol. 304:435-445(2000).
CC -I- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYZES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLURAMIDYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN.
CC -I- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminopurandioate + D-alanyl-D-alanine -> ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -I- PATHWAY: Peptidoglycan biosynthesis.
CC -I- SUBUNIT: MONOMER.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
CC -I- SIMILARITY: BELONGS TO THE MURCDF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15432; CA33473.1; -
DR EMBL: X55034; CA38863.1; -
DR EMBL: D10483; BAA01351.1; -
DR EMBL: U67891; AAC4657.1; -
DR EMBL: AE000118; AAC73197.1; -
DR EMBL: U67893; AAB60788.1; -
DR PIR: S04846; S04846.
DR PIR: S40596; S40596.
DR PDB: 1GG4; 20-DEC-00.
DR Ecogene: EG10622; murF.
DR InterPro: IPR000713; Mur_ligase.
DR InterPro: IPR004101; Mur_ligase_C.
DR Pfam: PF01225; Mur_ligase; 1.
DR TIGRfams: TIGR01143; murF; 1.
DR Pfam: PF02875; Mur_ligase_C; 1.
DR TIGRfams: TIGR01143; murF; 1.
KW Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KM ATP-binding; 3D-structure; Complete proteome.
FT NP BIND 107 113 ATP (POTENTIAL).
FT VARIANT 288 288 A->T (IN MURF2; TS MUTANT WITH LOW
FT ACTIVITY).
FT CONFLICT 61 61 G->A (IN REF. 1 AND 2).
FT FT 178 178 A->R (IN REF. 1 AND 2).
SQ SEQUENCE 452 AA; 47447 MW; B46E2E57BDBBC572 CRC64;
Query Match 27.7%; Score 62; DB 1; Length 452;
Best Local Similarity 42.3%; Pred. No. 0.59;
Matches 11; Conservative 7; Mismatches 4; Indels 4; Gaps 1;
OY 12 KSGRLNAPMPNGV---AENNMI 33
DB 200 KAKGEIFSGLPENGTAIMNADNMDL 225
RESULT 9
YMY9_YEAST STANDARD; PRT: 297 AA.
AC 003161.
DR 01-NOV-1997 (Rel. 35, Created)
DR 01-NOV-1997 (Rel. 35, Last sequence update)
DR 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 34.0 kDa protein in CTF13-YPK2 intergenic region.
GN YMR099C OR YMS543.06C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

```

OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxId=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: BELONGS TO THE UPF0010 FAMILY. SIGNIFICANTLY RELATED
CC TO METAROTASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z49807; CA98990.1; -
DR SGD: S0004705; YMR099C.
DR InterPro: IPR001823; Aldl_epimerase.
DR Pfam: PF01263; Aldose-epim; 1.
KW Hypothetical protein
FT ACT_SITE 159 159 POTENTIAL.
SQ SEQUENCE 297 AA; 33956 MW; 122BDC4608B9A32B CRC64;
Query Match 27.2%; Score 61; DB 1; Length 297;
Best Local Similarity 45.0%; Pred. No. 0.5;
Matches 9; Conservative 7; Mismatches 4; Indels 0; Gaps 0;
OY 21 MPENGAENNMIWFGYTOE 40
DB 79 LPQGLARNSTWERTGOTKE 98
RESULT 10
LM1_MOUSE STANDARD; PRT: 3084 AA.
AC P19137.
DR 01-NOV-1990 (Rel. 16, Created)
DR 01-NOV-1990 (Rel. 16, Last sequence update)
DR 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-1 chain precursor (Laminin A chain).
GN LM1 OR LAMA-1 OR LAMA.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89034134; PubMed=3182802;
RA Sasaki M., Kleinman H.K., Huber H., Deutzmann R., Yamada Y.;
RT "Laminin, a multidomain protein. The A chain has a unique globular
RT domain and homology with the basement membrane proteoglycan and the
RT laminin B chains."
RL Eur. J. Biol. Chem. 263:16536-16544(1988).
RN [2]
RP SEQUENCE OF 1-339 FROM N.A.
RX MEDLINE=88225080; PubMed=3267223;
RA Hartl L., Oberbauer I., Deutzmann R.;
RT "The N terminus of laminin A chain is homologous to the B chains."
RL Eur. J. Biochem. 173:629-635(1988).
RN [3]
RP SEQUENCE OF 2538-3084 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=89030693; PubMed=3181157;
RA Deutzmann R., Huber J., Schmeltz K.A., Oberbauer I., Hartl L.;
RT "Structural study of long arm fragments of laminin. Evidence for
RT repetitive C-terminal sequences in the A-chain, not present in the B-
RT chains."
RL Eur. J. Biochem. 177:35-45(1988).
CC -I- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting

```


FT DISULFID 1549 1560 BY SIMILARITY.
 FT DISULFID 1563 1563 INTERCHAIN (PROBABLE).
 FT DISULFID 1567 1567 INTERCHAIN (PROBABLE).
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 370 370 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 562 562 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 672 672 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 770 770 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 857 857 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 914 914 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 959 959 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 969 969 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1052 1052 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1344 1344 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1414 1414 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 25.4% Score 57; DB 1; Length 3084;
 Best Local Similarity 30.0%; Pred. No. 26;
 Matches 15; Conservative 6; Mismatches 15; Indels 14; Gaps 2;

QY 5 NRYRYALKSGR-----LNAPMPENGYAENND-----WINGYIQE 40
 DB 1260 NPEOYLINGRARKHYIMADAPAPENGVRDYEOMKEEFKYNYSVE 1309

RESULT 11
 TSH_DROME STANDARD; PRT; 993 AA.

AC P22265;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-AUG-1991 (Rel. 19, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Teashirt protein.

GN TSH.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91098655; PubMed=1846092;
 RA Fasano L., Roeder L., Core N., Alexandre E., Vola C., Jacq B.,
 RA Kerridge S.;
 RA "The gene teashirt is required for the development of Drosophila
 RT embryonic trunk segments and encodes a protein with widely spaced
 RT zinc finger motifs.";
 RL Cell 64:63-79(1991).
 RN [2]

RP POSSIBLE FUNCTION.
 RX MEDLINE=93083418; PubMed=1360402;
 RA Roeder L., Vola C., Kerridge S.;
 RA "The role of the teashirt gene in trunk segmental identity in
 RT Drosophila.";
 RL Development 115:1017-1033(1992).
 RN [3]

RP POSSIBLE FUNCTION.
 RX MEDLINE=95009555; PubMed=7925029;
 RA de Zulueta P., Alexandre E., Jacq B., Kerridge S.;
 RA "Homeotic complex and teashirt genes co-operate to establish trunk
 RT segmental identities in Drosophila.";
 RL Development 120:2287-2296(1994).

CC -I- FUNCTION: REQUIRED FOR THE DEVELOPMENT OF DROSOPHILA EMBRYONIC
 CC TRUNK SEGMENTS. REPRESS THE EXPRESSION OF HEAD HOMEOIC GENES.
 CC NECESSARY, IN COMBINATION WITH SCR, FOR THE FORMATION OF THE
 CC PROTHORACIC SEGMENT. MAY BE INVOLVED IN CHROMATIN STRUCTURE FOR
 CC MODULATION OF TRANSCRIPTION. NEGATIVELY CONTROL THE EXPRESSION OF
 CC MOD AND POSITIVELY THAT OF DTL AND OF ITS OWN EXPRESSION.

CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE TRUNK (FROM PS3

CC TO PS13).
 CC -I- DEVELOPMENTAL STAGE: EXPRESSED THROUGHOUT EMBRYONIC, LARVAL AND
 CC ADULT DEVELOPMENT. NOT MATERNALLY EXPRESSED.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.itsb-sib.ch/announce/>
 CC or send an email to license@itsb-sib.ch).

CC EMBL: M57496; AAA28983.1; -
 CC PIR: A38437; A38437.
 CC TRANSFAC: T00805; -
 CC FLYBASE: FBgn0003866; tsh.
 CC InterPro: IPR000822; Znf_C2H2.
 CC Pfam: PF00096; Zf-C2H2; 3.

DR SMART: SM00355; ZNF_C2H2; 3.
 DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE: PS50157; ZINC_FINGER_C2H2_2; 3.
 KW Developmental protein; Transcription regulation; Repressor; Activator;
 KW Zinc-finger; Metal-binding; Repeat; DNA-binding; Nuclear protein.

FT DOMAIN 104 136
 FT DOMAIN 175 183 ASP/GLU-RICH (ACIDIC).
 FT DOMAIN 354 557 ZINC FINGERS.
 FT ZN_FING 354 378 C2H2-TYPE.
 FT ZN_FING 466 490 C2H2-TYPE.
 FT ZN_FING 533 557 C2H2-TYPE.
 FT DOMAIN 104 107 POLY-ALA.
 FT DOMAIN 115 122 POLY-ALA.
 FT DOMAIN 175 180 POLY-GLU.
 FT DOMAIN 401 407 POLY-PRO.
 FT DOMAIN 830 834 POLY-ASN.
 SQ SEQUENCE 993 AA; 106206 MW; 2DF9C677AF68B6D1 CRC64;

Query Match 24.8% Score 55.5; DB 1; Length 993;
 Best Local Similarity 38.7%; Pred. No. 11;
 Matches 12; Conservative 5; Mismatches 5; Indels 9; Gaps 1;

QY 2 YKRYRYALKSGG-----RLNAPME 23
 DB 740 YQHYRTSSRSSECSAEARPLDAPPE 770

RESULT 12
 LYTB_STRPN STANDARD; PRT; 658 AA.

AC Q924P7;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative endo-beta-N-acetylglucosaminidase precursor (EC 3.2.1.96)
 DE (Murein hydrolase).
 GN LYTB OR SP0965.

OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-30.
 RC STRAIN=R6;
 RX MEDLINE=99195827; PubMed=10096093;
 RA Garcia P., Gonzalez M.P., Garcia E., Lopez R., Garcia J.L.,
 RT "Lyb, a novel pneumococcal murein hydrolase essential for cell
 RT separation.";
 RL Mol. Microbiol. 31:1275-1281(1999).
 RN [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=TRG4;
 RX MEDLINE=21357209; PubMed=11463916;
 RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,

RA Durkin A.S., Gwin M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umeyay L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzaple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguioi S., Dickenson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollinshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
CC -I- FUNCTION: PLAYS AN IMPORTANT ROLE IN CELL WALL DEGRADATION AND
CC CELL SEPARATION.
CC -I- CATALYTIC ACTIVITY: Endohydrolysis of the di-N-acetylchitobiosyl
CC unit in high-mannose glycopeptides and glycoproteins containing
CC the -[Man(GlcNAc)2]Asn-structure. One N-acetyl-D-glucosamine
CC residue remains attached to the protein; the rest of the
CC oligosaccharide is released intact.
CC -I- SIMILARITY: BELONGS TO FAMILY 73 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ010312; CA09078.1; -;
CC EMBL: AE007400; AK75086.1; -;
CC TIGR: SP0965; -;
CC InterPro: IPR002901; Amidase_4.
CC Pfam: PF01832; Amidase_4; 1.
CC SMART: SM00047; LY22; 1.
CC Signal: Hydrolyase; Cell wall; Complete proteome.
CC FT CHAIN 1 23
CC FT 24 658 PUTATIVE ENDO-BETA-N-
CC FT 336 336 ACETYLGLUCOSAMINIDASE.
FT 381 381 T -> M (IN REF. 1).
FT 384 384 E -> K (IN REF. 1).
FT 535 535 L -> P (IN REF. 1).
FT 580 580 F -> S (IN REF. 1).
SQ SEQUENCE 658 AA; 76469 MW; B625515006C9CB36 CRC64;

Query Match 24.6%; Score 55; DB 1; Length 658;
Best Local Similarity 37.5%; Pred. No. 8.3;
Matches 12; Conservative 5; Mismatches 5; Indels 10; Gaps 1;

OY 3 YKNRYRYALKSGRLNAPMPENVAENNDWIF 34
DB 213 FENGHYLYLKSGGYMA-----NEMIW 234
:::|||||:|:|:|
LMAL_HUMAN
ID LMAL_HUMAN STANDARD: PRT; 3075 AA.
AC P25391;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Laminin alpha-1 chain precursor (laminin A chain).
GN LAMA1 OR LAMA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-91333420; PubMed-1714537;
RA Haaparanta T., Uitto J., Ruoslahti E., Engvall E.;
RT "Molecular cloning of the cDNA encoding human laminin A chain.";
RL Matrix 11:151-160(1991).
RN [2]
RP SEQUENCE OF 1-2628 FROM N.A.

RX MEDLINE-91264789; PubMed-2049067;
RA Nissinen M., Vuolteenaho R., Boot-Handford R., Kallunki P.,
RA Tryggvason K.;
RT "Primary structure of the human laminin A chain. Limited expression
RT in human tissues.";
RL Biochem. J. 276:369-379(1991).
RN [3]
RP SEQUENCE OF 2397-3072 FROM N.A.
RX MEDLINE-89280632; PubMed-2733383;
RA Olsen D., Nagayoshi T., Fazio M., Peltonen J., Jaakkola S.,
RA Sandborn D., Sasaki T., Kuivaniemi H., Chu M.L., Deutzmann R.,
RA Timpi R., Uitto J.;
RT "Human laminin: cloning and sequence analysis of cDNAs encoding A, B1
RT and B2 chains, and expression of the corresponding genes in human
RT skin and cultured cells.";
RL Lab. Invest. 60:772-782(1989).
CC -I- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -I- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule
CC comprising one long and three short arms with globules at each
CC end.
CC THE ALPHA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ) AND
CC LAMININ-3 (S-LAMININ).
CC -I- SUBCELLULAR LOCATION: Extracellular.
CC -I- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -I- DOMAIN: THE ALPHA-HELIICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -I- DOMAIN: DOMAINS VI, IV AND G ARE GLOBALAR.
CC -I- SIMILARITY: CONTAINS 1 LAMININ N-TERMINAL DOMAIN (DOMAIN VI).
CC -I- SIMILARITY: CONTAINS 17 LAMININ EGF-LIKE DOMAINS.
CC -I- SIMILARITY: CONTAINS 2 LAMININ DOMAINS IV.
CC -I- SIMILARITY: CONTAINS 5 LAMININ G-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X58531; CA041418.1; -;
CC PIR: S1458; S1458.
CC HSP: 060675; 1000.
CC DR Genew: HGNC:6481; LAMA1.
CC MIM: 150320; -;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001886; LAMNT.
DR InterPro: IPR000034; Laminin_B.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR001791; Laminin_G.
DR Pfam: PF00052; laminin_B; 2.
DR Pfam: PF00053; laminin_EGF; 15.
DR Pfam: PF00054; laminin_G; 5.
DR Pfam: PF00055; laminin_Nterm; 1.
DR PRINTS: PR00011; EGF_LAMININ.
DR ProDom: PD002082; LAMNT; 1.
DR ProDom: PD003031; Laminin_B; 2.
DR SMART: SM00180; EGF_Lam; 14.
DR SMART: SM00001; EGF_Like; 1.
DR SMART: SM00281; Lamb; 2.
DR SMART: SM00282; Lamg; 5.
DR SMART: SM00136; Lamnt; 1.
DR PROSITE: PS00022; EGF_1; 11.
DR PROSITE: PS01186; EGF_2; 2.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 15.
DR PROSITE: PS50025; LAM_G-DOMAIN; 5.
KW Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M58002; AAA67325.1; -
 DR PIR: A38109; A38109
 DR InterPro: IPR002901; Amidase_4.
 DR InterPro: IPR002482; LysM.
 DR Pfam: PF01476; LysM; 5.
 DR SMART: PF01832; Amidase_4; 1.
 DR SMART: SM00047; Lys2; 1.
 DR SMART: SM00257; LysM; 5.
 KW Hydrolase; Glycosidase; Bacteriolytic enzyme; Cell wall;
 KW Cell division; Septation; Repeat; Signal.
 FT SIGNAL 1 53
 FT CHAIN 54 671 AUTOLYSIN.
 FT REPEAT 363 405 LYSM 1.
 FT REPEAT 431 473 LYSM 2.
 FT REPEAT 499 541 LYSM 3.
 FT REPEAT 567 609 LYSM 4.
 FT REPEAT 629 671 LYSM 5.
 SQ SEQUENCE 671 AA: 70442 MW: 34582686C6C1A4A33 CRC64;

Query Match 23.9%; Score 53.5; DB 1; Length 671;
 Best Local Similarity 37.5%; Pred. No. 14;
 Matches 13; Conservative 3; Mismatches 13; Indels 9; Gaps 1;

QY 8 YYALKSGRLNAPMPENGVAENN-----DWIFMGYT 38
 Db 566 YTIKSGDTLNKISAQFGVSVANLRSMNGIKGLIRFAGQT 605

RESULT 15
 RN12_YEAST
 ID RN12_YEAST STANDARD: PRT: 850 AA.
 AC P32843:
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE RNA12 protein.
 GN RNA12 OR PRP12 OR YMR302C OR YMR952.04C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 28583 / FL100;
 RX MEDLINE=92212295; PubMed=1557037;
 RA Liang S., Alksne L., Warner J.R., Lacroite F.;
 RT "RNA12", a gene of *Saccharomyces cerevisiae* involved in pre-rRNA
 RT maturation. Characterization of a temperature-sensitive mutant,
 RT cloning and sequencing of the gene.";
 RL Mol. Genet. 232:304-312(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Connor R., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (Apr-1995) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PLAYS A DISPENSABLE ROLE IN EARLY MATURATION OF
 CC PRE-RRNA. IT MAY PROVIDE A COFACTOR WHICH PLAYS A MINOR AND
 CC DISPENSABLE ROLE IN PRE-RRNA MATURATION. THIS PROTEIN MIGHT
 CC MIGRATE PROGRESSIVELY ALONG EACH PRE-RNA MOLECULE IN THE EARLY
 CC STEPS OF THE RNA MATURATION PATHWAY AND BE RELEASED ONCE THE
 CC MODIFICATION IT CATALYSES HAD BEEN COMPLETED.
 CC -!- SIMILARITY: TO S.POMBE SPBC83.05.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: S92205; AAB21991.1; -
 DR EMBL: Z49212; CA89135.1; -
 DR PIR: S20462; S20462.
 DR SGD: S0004917; PRP12.
 DR InterPro: IPR000504; RNA_rec_mot.
 DR Pfam: PF00076; rrm; 1.
 DR SMART: SM00360; RRM; 1.
 FT VARIANT 502 502
 FT N -> Y (IN RNA12-1; DOMINANT TS
 FT PHENOTYPE, EXHIBITS ABERRANT PROCESSING
 FT OF PRE-RRNA AND UNDERMETHYLATION OF
 FT NASCENT RRNA).
 SQ SEQUENCE 850 AA: 96688 MW: 3BEFD730AAD376C7 CRC64;

Query Match 23.9%; Score 53.5; DB 1; Length 850;
 Best Local Similarity 44.8%; Pred. No. 18;
 Matches 13; Conservative 3; Mismatches 12; Indels 1; Gaps 1;

QY 4 KNYRYALKS-GGRLNAPMPENGVAENNND 31
 Db 597 EDLYLNKKSKSGENVAKEPESEKETAENNND 625

Search completed: December 10, 2002, 10:54:44
 Job time : 7.6 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 10, 2002, 10:49:03 ; Search time 21.8 Seconds
(without alignments)
378.068 Million cell updates/sec

Title: US-09-142-970-5_COPY_1_40
Perfect score: 224
Sequence: 1 LYKKNRYVALKSGRLNAPMPENGVAENNDIMFGTQE 40

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues
Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP-organella:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP_vertebrate:*
14: SP_unclassified:*
15: SP_virus:*
16: SP_bacteriap:*
17: SP_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	210	93.8	993	2	09S6X3
2	210	93.8	996	2	030574
3	210	93.8	997	2	09S6X2
4	210	93.8	1552	2	09A106
5	198	88.4	996	2	057309
6	190	84.8	992	2	057035
7	190	84.8	992	2	09S6X5
8	190	84.8	992	2	09S6X4
9	190	84.8	992	2	030573
10	190	84.8	997	2	030575
11	190	84.8	1561	2	051169
12	190	84.8	1773	16	09JVB9
13	186	83.0	1815	16	09K0B4
14	102	45.5	1764	2	09J3T34
15	62	27.7	452	16	08ZRU6
16	62	27.7	452	16	08Z9H2

17	62	27.7	452	16	08X9Z1
18	57	25.4	189	17	026218
19	57	25.4	1367	5	09RYU0
20	56.5	25.2	435	5	09NFT4
21	56	25.0	217	17	096XX6
22	55.5	24.8	766	5	018673
23	55.5	24.8	948	5	09V900
24	55	24.6	161	11	08VEA3
25	55	24.6	574	16	08XEC2
26	55	24.6	614	2	09AHT8
27	54	24.1	283	9	09AFB0
28	54	24.1	350	2	043983
29	54	24.1	353	2	032433
30	54	24.1	473	16	08RDT2
31	54	24.1	1363	12	08Y436
32	54	24.1	2146	13	09PUM3
33	54	24.1	2157	13	09PW44
34	54	24.1	2171	13	09PW46
35	54	24.1	2182	13	09PUM5
36	54	24.1	2321	13	09PUM4
37	54	24.1	2332	13	09PW45
38	54	24.1	2346	13	09PW47
39	54	24.1	2357	13	09PUM6
40	53.5	23.9	455	5	024464
41	53.5	23.9	459	5	09NGW6
42	53.5	23.9	459	5	09VW46
43	53.5	23.9	463	5	09NBK5
44	53.5	23.9	660	2	057443
45	53	23.7	162	2	09R612

ALIGNMENTS

RESULT 1	PRELIMINARY;	PRT;	993 AA.
09S6X3	09S6X3		
AC	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	IgA1 protease (Fragment).		
GN	IGA.		
OS	Neisseria meningitidis.		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=4877;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Z4400;		
RA	MEDLINE=98010345; PubMed=9350862;		
RA	Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,		
RA	Achman M.;		
RT	"Clonal descent and microevolution of Neisseria meningitidis during 30		
RT	years of epidemic spread."		
RI	Mol. Microbiol. 25:1047-1064(1997).		
RL	EMBL: AF012208; AAC45791.2;		
DR	MEROPS, S06.001;		
DR	InterPro: IPR002195; Dihydroorotase.		
DR	InterPro: IPR000710; IGA_S6.		
DR	InterPro: IPR004899; Pertactin_sup.		
DR	Pfam: PF02395; IGA1; 1.		
DR	Pfam: PF03212; Pertactin; 1.		
DR	PRINTS: PR00921; IGASERPTASE.		
DR	PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.		
KW	Protease.		
FT	NON_TER	1	
FT	NON_TER	993	
SO	SEQUENCE	993 AA.	109441 MW; 109FAA2EF88AC3C6 CRC64;
Query Match	93.8%;	Score 210;	DB 2; Length 993;
Best Local Similarity	90.0%;	Pred. No. 1.1e-19;	
Matches	36;	Conservative	2; Mismatches 2; Indels 0; Gaps 0;

```
OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNNDWIFMGYTOE 40
DB 558 LYKKNRYRYALKSGGSVNAPEENGVTENNNDWVFMGYTOE 597

RESULT 2
O30574
ID 030574 PRELIMINARY; PRT; 996 AA.
AC 030574;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN IGA1 protease (Fragment).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24024;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012206; AAC45789.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1
FT NON_TER 996
SQ SEQUENCE 996 AA; 109688 MW; B3ABDEFE54C337D9 CRC64;

Query Match 93.8%; Score 210; DB 2; Length 996;
Best Local Similarity 90.0%; Pred. No. 1,1e-19;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNNDWIFMGYTOE 40
DB 557 LYKKNRYRYALKSGGSVNAPEENGVTENNNDWVFMGYTOE 596

RESULT 3
O9S6X2
ID 09S6X2 PRELIMINARY; PRT; 997 AA.
AC 09S6X2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN IGA1 protease (Fragment).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24081;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
  Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
  years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012209; AAC45792.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
```

```
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1
FT NON_TER 997
SQ SEQUENCE 997 AA; 109812 MW; 06F2E361E7E202E0 CRC64;

Query Match 93.8%; Score 210; DB 2; Length 997;
Best Local Similarity 90.0%; Pred. No. 1,1e-19;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNNDWIFMGYTOE 40
DB 558 LYKKNRYRYALKSGGSVNAPEENGVTENNNDWVFMGYTOE 597

RESULT 4
O9A1U6
ID 09A1U6 PRELIMINARY; PRT; 1552 AA.
AC 09A1U6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN IGA1 protease.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NMB;
RA Vitovski S., Sayers J.R.;
RT "Degenerate specificity of Neisseria meningitidis Iga1 protease.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF235032; AKR15023.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydroorotase.
DR InterPro: IPR000710; Iga_S6.
DR InterPro: IPR004899; Pertact_sup.
DR Pfam: PF02395; IGA1; 1.
DR Pfam: PF03212; Pertactin; 2.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
SQ SEQUENCE 1552 AA; 171205 MW; E49D6657B120CEFF CRC64;

Query Match 93.8%; Score 210; DB 2; Length 1552;
Best Local Similarity 90.0%; Pred. No. 1,1e-19;
Matches 36; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNNDWIFMGYTOE 40
DB 574 LYKKNRYRYALKSGGSVNAPEENGVTENNNDWVFMGYTOE 613

RESULT 5
O57309
ID 057309 PRELIMINARY; PRT; 996 AA.
AC 057309;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
GN IGA1 protease precursor (Fragment).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23524;
```

```

RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
RN [2]
RP SEQUENCE OF 37-532 FROM N.A.
RC STRAIN=ETH2;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
EMBL AF012207; AAC45790.2; -.
DR EMBL; X82469; CAA57852.1; -.
DR EMBL; X82468; CAA57851.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; Iga1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1 1
FT NON_TER 996 996
SO SEQUENCE 996 AA; 109717 MW; AB39FB98A4F986 CRC64;

Query Match
Best Local Similarity 84.8%; Score 198; DB 2; Length 992;
Matches 34; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 LYKKNRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 557 LYKKNRYALKSGGSVNAPEMNGOTENNNDWIFMGYKQE 596

RESULT 6
ID 057035 PRELIMINARY; PRT; 992 AA.
AC 057035;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease precursor (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B40;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Mueller K., Sella A., Wang J.F.,
RA del Valle J., Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
RN [2]
RP SEQUENCE OF 37-532 FROM N.A.
RC STRAIN=HF46;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding Iga1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
EMBL AF012211; AAC45794.2; -.
DR EMBL; X82480; CAA57863.1; -.
DR EMBL; X82475; CAA57858.1; -.

```

```

DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; Iga1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SO SEQUENCE 992 AA; 109228 MW; 3677DDEACE6D9F69 CRC64;

Query Match
Best Local Similarity 84.8%; Score 190; DB 2; Length 992;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYKKNRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 557 LYKKNRYALKSGGSVNAPEMNGOTENNNDWIFMGSTOE 596

RESULT 7
ID 09S6X5 PRELIMINARY; PRT; 992 AA.
AC 09S6X5;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Iga1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Sella A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
RT years of epidemic spread.";
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL; AF012203; AAC45786.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000710; Iga_S6.
DR InterPro; IPR004899; Pertactin_sup.
DR Pfam; PF02395; Iga1; 1.
DR Pfam; PF03212; Pertactin; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KM Protease.
FT NON_TER 1 1
FT NON_TER 992 992
SO SEQUENCE 992 AA; 109239 MW; 306FF451B7213D49 CRC64;

Query Match
Best Local Similarity 84.8%; Score 190; DB 2; Length 992;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 LYKKNRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 557 LYKKNRYALKSGGSVNAPEMNGOTENNNDWIFMGSTOE 596

RESULT 8
ID 09S6X4 PRELIMINARY; PRT; 992 AA.
AC 09S6X4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

```

```
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23906;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012204; AAC45787.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydrocoarataase.
DR InterPro: IPR000710; IGA_S6.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109239 MW; 9ED0A41019917CE3 CRC64;

Query Match 84.8%; Score 190; DB 2; Length 992;
Best Local Similarity 85.0%; Pred. No. 5.5e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 557 LYKKNRYRYALKSGGSYNAPMPENGOTENNNDWILMGSTOE 596
|||||
OY 030573 PRELIMINARY; PRT; 992 AA.
AC 030573;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23910;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012205; AAC45788.1; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydrocoarataase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin-sup.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 992
SQ SEQUENCE 992 AA; 109255 MW; EF7CA563BC7D7ECB CRC64;

Query Match 84.8%; Score 190; DB 2; Length 992;
```

```
Best Local Similarity 85.0%; Pred. No. 5.5e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 557 LYKKNRYRYALKSGGSYNAPMPENGOTENNNDWILMGSTOE 596
|||||
OY 030575 PRELIMINARY; PRT; 997 AA.
AC 030575;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE IGA1 protease (Fragment).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=24099;
RX MEDLINE=98010345; PubMed=9350862;
RA Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA Achman M.;
RT "Clonal descent and microevolution of Neisseria meningitidis during 30
years of epidemic spread."
RL Mol. Microbiol. 25:1047-1064(1997).
DR EMBL: AF012210; AAC45793.2; -.
DR MEROPS: S06.001; -.
DR InterPro: IPR002195; Dihydrocoarataase.
DR InterPro: IPR000710; IGA_S6.
DR InterPro: IPR004899; Pertactin-sup.
DR Pfam: PF02395; IGA1; 1.
DR PRINTS: PR00921; IGASERPTASE.
DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW Protease.
FT NON_TER 1
FT NON_TER 997
SQ SEQUENCE 997 AA; 110152 MW; 4BE40DE594E5E2BE CRC64;

Query Match 84.8%; Score 190; DB 2; Length 997;
Best Local Similarity 85.0%; Pred. No. 5.5e-17;
Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYRYALKSGGRINAPMPENGVAENNDWIFMGYTOE 40
DB 558 LYKKNRYRYALKSGGSYNAPMPENGOTENNNDWILMGSTOE 597
|||||
OY 051169 PRELIMINARY; PRT; 1561 AA.
AC 051169;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
DE IGA1 protease.
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HR13;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IGA1 protease
in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
Influenzae."
RL Mol. Microbiol. 15:495-506(1995).
```

DR EMBL: X82474; CAA57857.1; -
 DR MEROPS: S06.001; -
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertactin_sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 2.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Protease.
 KW
 SQ SEQUENCE 1561 AA; 171849 MW; 1C96E291A0001705 CRC64;

Query Match 84.8%; Score 190; DB 2; Length 1561;
 Best local Similarity 85.0%; Pred. No. 9, 4e-17;
 Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAMPENGVAENNDWIFMGYTOE 40
 |||
 DB 584 LYKKNRYVALKSGGSYNAPMENGOTENDWILMGSTOE 623

RESULT 12
 O9JVB9 PRELIMINARY; PRT; 1773 AA.

AC O9JVB9;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE IGA1 protease (EC 3.4.21.7).
 GN IGA OR NMA0905.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=65699;

RN [1]
 SEQUENCE FROM N.A.
 RC STRAIN=24291 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=2022556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Fellwell T., Hamlin N., Holtroyd S.,
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajadream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrell B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491.";
 RL Nature 404:502-506(2000).

DR EMBL: AL162754; CAB84182.1; -
 DR MEROPS: S06.001; -
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertactin_sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 2.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Protease; Hydrolyase; Complete proteome.
 SQ SEQUENCE 1773 AA; 196350 MW; CAC19E7313D6CE1 CRC64;

Query Match 84.8%; Score 190; DB 16; Length 1773;
 Best local Similarity 85.0%; Pred. No. 1, 1e-16;
 Matches 34; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAMPENGVAENNDWIFMGYTOE 40
 |||
 DB 573 LYKKNRYVALKSGGSYNAPMENGOTENDWILMGSTOE 612

RESULT 13
 O9KOB4 PRELIMINARY; PRT; 1815 AA.

AC O9KOB4;
 DT 01-OCT-2000 (TREMblrel. 15, Created)
 DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)

DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE IGA-specific serine endopeptidase.
 GN NMB0700.
 OS Neisseria meningitidis (serogroup B).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=491;
 RX [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=MC58 / SEROGROUP B;
 RX MEDLINE=20175755; PubMed=10710307;
 RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
 RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
 RA Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K.,
 RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
 RA Mason T., Ciecko A., Parksey D.S., Blair E., Citron H., Clark E.B.,
 RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
 RA Gill J., Scarlato V., Maignani V., Pizsa M., Grandi G., Sun L.,
 RA Smith H.O., Fraser C.M., Moxon E.R., Rappoli R., Venter J.C.;
 RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
 MC58.";
 RL Science 287:1809-1815(2000).
 DR EMBL: AE002424; AAF4117.1; -
 DR MEROPS: S06.001; -
 DR TIGR: NMB0700; -
 DR InterPro: IPR002195; Dihydroorotase.
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertactin_sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR PRINTS: PR00921; IGASERPTASE.
 DR PROSITE: PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
 DR Complete proteome.
 KW
 SQ SEQUENCE 1815 AA; 201077 MW; 2259D4D71762C57F CRC64;

Query Match 83.0%; Score 186; DB 16; Length 1815;
 Best local Similarity 82.5%; Pred. No. 3, 9e-16;
 Matches 33; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 1 LYKKNRYVALKSGGRNAMPENGVAENNDWIFMGYTOE 40
 |||
 DB 573 LYKKNRYVALKSGGSYNAPMENGOTENDWILMGSTOE 612

RESULT 14
 O93T34 PRELIMINARY; PRT; 1764 AA.

AC O93T34;
 DT 01-DEC-2001 (TREMblrel. 19, Created)
 DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
 DE IGA1 protease type 2.
 GN IGA1.

OS Haemophilus aegyptius.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 NCBI_TaxID=725;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=F3031;
 RA MCGillivray G., Actis L.A.;
 RT "Iga protease from H. aegyptius F3031.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF369807; AAK56925.1; -
 DR MEROPS: S06.001; -
 DR InterPro: IPR000710; IGA_S6.
 DR InterPro: IPR004899; Pertactin_sup.
 DR Pfam: PF02395; IGA1; 1.
 DR Pfam: PF03212; Pertactin; 1.
 DR Protease.

KW
 SQ SEQUENCE 1764 AA; 193863 MW; EC583CDB81DBEC CRC64;

Query Match 45.5%; Score 102; DB 2; Length 1764;
 Best local Similarity 44.7%; Pred. No. 9, 5e-05;

